92
2ej

UNIVERSIDAD NACIONAL AUTONOMA
DE MÉXICO

FACULTAD DE CIENCIAS

APUNTES DE APOYO PARA EL NUEVO PROGRAMA DE ESTADÍSTICA Y PROBABILIDAD I DE LA ESCUELA NACIONAL COLEGIO DE CIENCIAS Y HUMANIDADES

T E S I S

QUE PARA OBTENER EL TÍTULO DE ACTUARIO

P R E S E N T A :

ARMANDO TERRÉS SANDOVAL

M. EN C. JOSÉ ANTONIO FLORES DÍAZ

FACULTAD DE CIENCIAS
U.N.A.M.

1999

VERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

M. en C. Virginia Abrín Batule
Jefe de la División de Estudios Profesionales de la
Facultad de Ciencias
P r e s e n t e

Comunicamos a usted que hemos revisado el trabajo de Tesis:

Apuntes de Apoyo para el nuevo programa de Estadística y Probabilidad I
de la Escuela Nacional Colegio de Ciencias y Humanidades.

realizado por ARMANDO TERRES SANDOVAL

con número de cuenta 7126639-3 , pasante de la carrera de Actuario

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis Propietario	M. en C. José Antonio Flores Díaz.	
Propietario	M. en A.P. María del Pilar Alonso Reyes.	
Propietario	Act. María Guadalupe Tzintzún Cervantes	
Suplente	Fís. Mat. José Luis Macías Avila.	
Suplente	Act. Jaime Vázquez Alamilla.	

Consejo Departamental de Matemáticas

M. en A.P. ~~Ma. del Pilar~~ Alonso Reyes.

A Dios nuestro Señor
por haberme dado la
vida y el animo para
lograr terminar mi
carrera

A mis padres que Dios tenga
en su gloria y que siempre me
brindaron su ayuda, me dieron
su amor y a los que desearia
tener conmigo en este momento

A mi esposa que es mi compañera
eterna y que ademas de darme su
comprensión me dio a la hija más
hermosa de la tierra.

A mis hermanos con los cuales
conviví en los momentos más
difíciles de mi carrera y a los
cuales les doy las gracias por
ser mis hermanos.

Al Sr. Ing. Genaro Briz Plata
que me ayudó a tomar la decisión
adecuada, sin la cual esto
no se hubiera realizado

PROGRAMA

4.- PROBABILIDAD

4.1 Probabilidad Clásica y Frecuencial.

Conceptos y definición.

4.2 Leyes de la Probabilidad

4.2.1 Evento Seguro y Evento Imposible.

4.2.2 Unión, Intersección y Complemento de Eventos

4.2.3 Probabilidad Condicional y Eventos Independientes

4.3 Técnicas del Conteo.

4.3.1 Teorema Fundamental de la Aritmética, Cálculo de Combinaciones y Diagramas de Árbol.

4.4 Variable Aleatoria.

4.4.1 Esperanza Matemática.

4.4.2 Desviación Estándar.

4.4.3 Varianza.

5.- FUNCIÓN DE DISTRIBUCIÓN.

5.1 Tipo de Distribución de Probabilidad.

5.1.1 Uniforme.
5.1.2 Bernulli.
5.1.3 Binomial.
5.1.4 Poisson
5.1.5 Normal.

5.2 Estudio de la Distribución Binomial, como ejemplo de Distribución de Pro-

1.- INTRODUCCIÓN

La Universidad Nacional Autonoma de Mexico como la maxima casa de estudios de nuestro pais, ha tenido que responder a los cambios que le demanda la modernizacion y el progreso, el esfuerzo de superacion ha estado de manifiesto en cada uno de los que pertenecemos y amamos a esta gran institucion educativa

Es por tal motivo que surge la inquietud de elaborar este trabajo cuyo principal objetivo es proporcionar, en la medida de sus posibilidades, una herramienta util para profesores y alumnos de La Escuela Nacional Colegio de Ciencias y Humanidades, ya que se apega al nuevo programa de Estadistica y Probabilidad I del plan actual La presentacion sera manejada como apuntes y seran auxiliados con diagramas, dibujos y cuadros sinopticos con el objeto de lograr una mejor comprension de esta materia

La informacion fue elegida con base a la experiencia docente adquirida y con el correspondiente fundamento bibliografico, asi como, varias horas de investigacion (por lo que se refiere a la reseña historica) Con respecto a la experiencia se observo que el alumno requiere de una motivacion constante y un mejor trato por parte del profesor, hacia la enseñanza de las matematicas Debido a ello, se pretende enfatizar en la estructura clara y explicita de los enunciados a tratar

De esta forma, actividad por actividad se hacen intentos por concientizar al alumno, de que es muy importante instalar el razonamiento mas no la memorizacion de un procedimiento

Dentro de esta nueva etapa que vive el Colegio de Ciencias y Humanidades, se da inicio a un nuevo proceso de enseñanza-aprendizaje donde educar es la parte fundamental , es aqui donde cualquier trabajo que se desarrolle en pro de un esfuerzo para ayudar a la mejor y mas rapida comprension de los cambios en los programas, tendra alguna justificacion Asimismo el trabajo esta desarrollado en seis capitulos, los capitulos uno y dos pretenden proporcional al lector los aspectos introductorios y una breve reseña historica del Colegio, donde se trata , ademas algunos causas que *originaron los cambios de programas En los capítulos tres, cuatro y cinco, se desarrollan los apuntes, basados en el programa de la materia que se impartirá a partir del año escolar 1998-1999. Cabe destacar, que en algunos puntos de estos apuntes, se llevarán a cabo algunos cambios de forma y no de fondo Por ejemplo, en el capítulo tres cuando se habla de cuartiles, las fórmulas son tratadas de forma un poco diferente, lo que también sucede en técnicas del conteo, tratadas en el capítulo cuatro.*

Se sabe que muchas y variadas nomenclaturas existen al respecto, sin embargo estos apuntes no pretenden competir con ellas, más bien unirse al tremendo esfuerzo que significa la enseñanza a nivel medio superior.

2.- ANTECEDENTES

El Colegio de Ciencias y Humanidades, ahora Escuela Nacional Colegio de Ciencias y Humanidades, desde su creación y hasta la fecha ha experimentado cambios apegados a los requerimientos de una poblacion joven, como es la mexicana, que demanda más y mejor educación a nivel medio superior

Uno de los problemas a los que se ha enfrentado el Colegio, es la elaboración de programas de estudio, tarea mucho muy importante en el difícil trabajo de la docencia

Esta tarea, permite concretar las diversas concepciones teóricas e ideologicas que sobre el acto educativo sustentan las personas que integran la institucion y esta debe ser realizada por los docentes responsables de cada unidad de enseñanza, ya que el programa es la herramienta fundamental del trabajo que realiza el docente y está intimamente relacionado con los problemas y con la intencionalidad que caracteriza a la práctica docente La elaboración de los programas de estud. proporciona una visión más profunda de la problemática que se afronta en el proceso de enseñanza-aprendizaje de un curso específico.

La participación del profesorado en la realizacion de los programas de estudio de los cursos que se imparten, refuerza la idea de la formación didáctica de un profesor, debe centrarse en el aprendizaje de técnicas de enseñanza y así mismo, en el análisis de la disciplina, la orientación pedagógica e ideologica , etcétera. Es por ello que se puede sugerir, que los profesores participen en la elaboracion de los mismos, y esto sea cada vez mayor, ya que son ellos los que al enfrentarse al alumno podrán proporcionar varias y valiosas ideas que fortalezcan a los encargados de esta difícil tarea

2.1.- RESEÑA HISTÓRICA

El objeto de presentar una reseña histórica del Colegio, es con el fin de que toda aquella persona que desee leer este trabajo, tenga antecedentes suficientes para comprender la necesidad de un cambio inevitable

Desde mayo de 1970, año en que se inició el Rectorado del Doctor Pablo Gonzales Casanova, comenzaron a realizarse estudios tendientes a reformar substancialmente la estructura y la metodología de la enseñanza en los tres niveles que cubre la Universidad Nacional Autonoma de México y que son Educacion Media Superior, Licenciatura y el Posgrado

Las labores iniciales fueron confiadas a un grupo no menor de ochenta destacados universitarios encabezados por el Doctor Roger Díaz de Cossio, Cordinador de Ciencias en aquella epoca

CREACIÓN DEL COLEGIO DE CIENCIAS Y HUMANIDADES

Son múltiples los factores que influyeron en la creacion del Colegio, éste no surgió como un hecho aislado, sino que formo parte de un proyecto mas amplio que estaba condicionado por una realidad histórica concreta La reforma educativa que se penso, buscaba combatir el problema de la desercion, aumentar la posibilidad de acceso a la enseñanza a grupos marginados, utilizando tecnicas pedagogicas modernas Es en este contexto, que surge la creacion del Colegio de Ciencias y Humanidades como parte de un proyecto tendiente a resolver, entre otras cosas, la renovación de las estructuras y los metodos educativos asi como, la creciente demanda de educación provocada por el acelerado crecimiento demografico de la poblacion Es asi, que el 26 de enero de 1971, el Consejo Universitario aprobó por unanimidad la creacion del Colegio de Ciencias y Humanidades, y el 12 de abril del mismo año, se *iniciaron la clases en los planteles Azcapotzalco, Vallejo y Naucalpan, proporcionando educacion media superior a 15,000 alumnos y un año mas tarde se les incorporan los planteles oriente y Sur con 10,000 alumnos mas. El Doctor Pablo Gonzalez Casanova, Rector de la máxima casa de estudios, con motivo de este acontecimiento que causaba una transformacion histórica en la vida educativa de la Universidad, señalo las bases fundamentales de la nueva institucion y que hoy pueden exponerse como los objetivos generales del CCH y que son.*

- *Crear un organo permanente de innovacion de la Universidad, capaz de realizar funciones distintas sin tener que cambiar toda la estructura universitaria, adaptando l sistema a los cambios y necesidades de la propia Universidad y del pais*
- *Preparar jovenes para cursar estudios que vinculen las humanidades, las ciencias y las técnicas, en el nivel de bachillerato de licenciatura, de maestria y de doctorado*
- *Proporcionar nuevas oportunidades de estudio acordes con el desarrollo de las ciencias y las humanidades en el siglo XX y hacer flexibles los sistemas de enseñanza para formar especialistas y profesionales que puedan adaptarse a un mundo cambiante en el terreno de la ciencia, la tecnica y las estructuras sociales y culturales.*
- *Intensificar la interdisciplina entre especialistas, escuelas, Facultades, centros e institutos de investigacion de la Universidad*
- *Promover el mejor aprovechamiento de los recursos humanos y tecnicos de la Universidad.*

Al crearse el Colegio, se presento al Consejo Universitario un esquema matricial del plan de estudios con una descripcion de las cuatro areas académicas de estudios, las 20 asignaturas obligatorias de los primeros cuatro semestres y las 44 materias de quinto y sexto semestres, con las cuales se podria hacer multiplicidad de opciones para que los bachilleres de este sistema cumplieran con las 33 asignaturas de dicho plan, mismo que comprenderia un idioma extranjero y adiestramiento técnico, *considerado como opcional. En abril de 1971, entonces y como guía para cumplir con el plan, se entregó a los profesores un temario que abordaba la asignatura que debía impartir.*

Posteriormente los maestros recibieron la Guía de Profesor del CCH. Con la información acerca de los objetivos, organigrama, plan de estudios, metodología y evaluación de aprendizajes en el CCH. Así como, un folleto de información académica del Colegio que incluía las bases pedagógicas del CCH, programas, evaluaciones y funciones de los jefes de área y maestros.

Por lo que toca a los estudiantes, aquel que lograra cubrir el total de los créditos podría seguir cualquier carrera de la Universidad o cualquiera de las combinaciones de carreras interdisciplinarias que establezcan el Colegio a nivel licenciatura. Se extendería diploma de bachiller a los que hayan cubierto dicho requisito.

Los alumnos podrían, sin asistir a clases acreditar los cursos de lenguas extranjeras mediante un examen en que demostrase su capacidad de traducción y comprensión del inglés o francés.

En los laboratorios se planteó que los alumnos construyeran algunos de los aparatos de observación y que los aplicaran.

Permanentemente el Colegio revisaría y en su caso, actualizaría el plan de estudios. Los programas deberían ser publicados anualmente.

Cada plantel de la Unidad Académica organizaría conferencias destinadas a explicar el plan de estudios y sus reglas de aplicación.

En cada plantel de la Unidad Académica, debería haber una planta de profesores de carrera, de asignatura y de ayudantes.

CRONOLOGÍA

1971

El 26 de enero, el Consejo Universitario aprueba la creación del Colegio de Ciencias y Humanidades y el 12 de abril se inician los cursos con 15,000 alumnos en tres planteles, Azcapotzalco, Naucalpan y Vallejo, con una planta de 450 profesores, seleccionados por concurso de entre 2,057 participantes que acudieron a la convocatoria de la Universidad.

1972

Se abren los planteles Oriente y Sur. La población escolar asciende a 40,000 estudiantes y la planta docente a 900. Se inicia la operación del programa de Opciones Técnicas y las actividades de Educación Física. Las academias de profesores tienen un peso importante en la selección de profesores y en la revisión de apoyo didáctico.

1973

La población estudiantil alcanza 65,000 alumnos La planta docente llega a 1,350 profesores y en los planteles de la norte concluye la tercera etapa de la construcción. Se publican las primeras guías de estudio para los exámenes extraordinarios.

1974

Se construye la Dirección de la Unidad Académica y las unidades escolar y administrativa, se impulsan después de una revisión y con un nuevo modelo, las Operaciones Técnicas y se aprueba su reglamento y su Consejo Academico, se inicia *el proceso de regulación y definitividad académica de profesores con la aprobación del acuerdo número 61 del Consejo Universitario. Egresa la primera generación de estudiantes de los planteles Azcapotzalco, Naucalpan y Vallejo, se crea la Secretaría de Planeación y se funda la Gaceta del CCH, órgano oficial del Colegio. Se promueven, desde la Dirección del Bachillerato, las actividades de superación de la academia de los profesores y los eventos propiciados por ellos en beneficio de la formación de los alumnos.*

1975

Se integra la primera Comisión Dictaminadora, se desarrollan los programas para instalar el profesorado de carrera de enseñanza superior, se crea la Secretaría Auxiliar de Planeación del CCH.

1976

El Consejo del Colegio aprueba las primeras definitividades para 523 profesores de Asignatura; El consejo Universitario aprueba los instructivos que dan base al programa de profesorado de Carrera de Enseñanza Media Superior y aprueba el proyecto de la Unidad Académica de los Ciclos Profesionales y de Posgrado. Se trabaja en cinco proyectos de Licenciatura, Maestría y Doctorado, el Consejo aprueba lo relativo a los programas de complementación y regularización académica, en agosto, es sometido al Consejo Universitario un proyecto de reglamento para la Unidad Académica del Bachillerato, para estas fechas, la planta docente del Colegio se ha destacado por su participación en los programas de formación y de actualización generados por la Universidad. Los cursos de didáctica programados a partir de la experiencia del Colegio en 1972 y 1973 alcanzan difusión a nivel nacional.

1977

Se crean en los planteles las comisiones docentes, se formaliza el Seminario Académico del Bachillerato como órgano auxiliar de la Junta de Directores, con la participación de los secretarios y docentes de los planteles, se aprueba el instructivo para la asignación de horarios, se completa la reglamentación del bachillerato, se construyen los edificios que puedan sustentar los programas de docencia y del profesorado de carrera en los planteles del bachillerato.

1978

Se instala el primer Consejo Académico por área y se ratifican las primeras plazas de profesores de carrera de enseñanza media superior, se intenta el aprovechamiento de los aprovechamiento de los materiales elaborados para el programa de autoaprendizaje, en apoyo a los grupos regulares, a estas fechas han egresado por esa vía 61 de los 156 alumnos registrados en el programa, los Consejos Académicos por área, seis en total, con los de Opciones Técnicas e Idiomas, tienen como función primordial la de auxiliar al Consejo del Colegio en lo que se refiere a la orientación y evaluación del trabajo de docencia, principalmente el que realizan los profesores de carrera y complementación académica, estos órganos están integrados por representantes de los profesores y las autoridades

1979

El Consejo del Colegio, aprueba los criterios de promoción de profesores de asignatura, y algunos profesores acceden a la categoría de asignatura B, se inicia la transformación histórica en la vida educativa de la Universidad, señalando las bases *fundamentales de la nueva institución y que hoy pueden exponerse como los objetivos generales del Colegio de Ciencias y Humanidades.*

Posteriormente los maestros recibieron la Guía de Profesor del CCH Con la informacion acerca de los objetivos, organigrama, plan de estudios, metodología y evaluación de aprendizajes en el CCH. Así como, un folleto de información académica del Colegio que incluía las bases pedagógicas del CCH, programas, evaluaciones y funciones de los jefes de área y maestros.

Por lo que toca a los estudiantes, aquel que lograrà cubrir el total de los créditos podría seguir cualquier carrera de la Universidad o cualquiera de las combinaciones de carreras interdisciplinarias que establezcan el Colegio a nivel licenciatura Se extendería diploma de bachiller a los que hayan cubierto dicho requisito.

1980

El Coordinador del Colegio, Lic. David Pantoja Morán, invita a los consejeros académicos, las comisiones dictaminadoras , las coordinaciones de área y a la comunidad académica del bachillerato, a un proceso de revisión del contenido y orientación de los programas de estudio, se establecen, en esa ocasión, los modelos de documentos que han de configurar los programas del bachillerato, el Colegio cuenta ya con aproximadamente 76,000 alumnos inscritos y una planta de 1,874 profesores de asignatura A , 54 de asignatura B, 12 de carrera de enseñanza media superior, 4 de carrera y 120 en el Programa de Complementación Académica

1981

Para conmemorar el Décimo Aniversario del Colegio, se activan de manera significativa, los eventos en que participan los maestros y alumnos del bachillerato *como el Primer Foro Nacional de Investigación en el Proceso de Enseñanza-Aprendizaje, el Congreso Nacional de Enseñanza de la Biología, el Congreso Nacional de Química y el Metropolitano de Enseñanza de las Matemáticas, así como, los concursos Inter-CCH de física.*

1982

Con ocasión del Simposio Internacional sobre el Bachillerato los profesores de la unidad generan más de 1,500 páginas de comentarios al documento base, que se recogen en seis volúmenes; En el Foro-Muestra de trabajos de profesores de carrera y complementación se dan a conocer cerca de 100 trabajos producidos en apoyo a la docencia.

1983

Se revisa la orientación del programa de formación de profesores y se inician acciones tendientes a fomentar la titulación de los maestros; Se impulsa decisivamente el trabajo de los consejos académicos y se les invita a asumir plenamente las funciones que en la publicación del suplemento de la Gaceta ECI (Experiencias, Comentarios e Información) dirigido a los profesores, se inician los trabajos de revisión de los Criterios Normativos del programa de apoyo a la docencia, se amplían las acciones de carácter extracurricular para los alumnos, se formaliza el proceso de planeación y evaluación en el Colegio y entre sus primeros resultados se presenta del programa universitario de evaluación la información detallada sobre el bachillerato, en una exposición celebrada en el Plantel Vallejo, ante el Rector de la Universidad y el Colegio de Directores de Facultades y Escuelas.

1984

Se aprueban los criterios normativos para el trabajo de apoyo a la docencia que simplifican los procesos de evaluación, posibilitan en mayor grado el trabajo colectivo, diversifican las actividades a realizar homogéneamente los periodos de trabajo y la socialización de los resultados del mismo, mediante el establecimiento de los seminarios por área, durante el periodo interanual, entre otras actividades, se realiza el Encuentro sobre las perspectivas del bachillerato del Colegio de Ciencias y Humanidades, en cuya ponencia institucional se encuentra no sólo información detallada sobre el estado de la unidad (en sus aspectos institucionales, alumnos, profesores, acción educativa y proyecto), sino planteamientos optativos cuestionamientos hechos desde la autoridad del Colegio, se hace la exposición " El libro de CCH " con una exhibición de más de 600 títulos producidos por profesores del bachillerato. Es el Seminario por área, formalizado por el Consejo del Colegio, en los Criterios Normativos, la reunión en que al menos anualmente los profesores de carrera, de complementación y de asignatura participan en el análisis de las necesidades y problemas académicos de su área-plantel, conocen y comentan la orientación y resultados del trabajo del profesorado de carrera en apoyo a la docencia, estas reuniones son indispensables para la formulación de los planes anuales de trabajo de las áreas, cuya elaboración correspondiente a los consejos académicos por área.

1987

Se efectúa la ceremonia de reconocimiento de alumnos sobresalientes del Colegio, se realiza la primera muestra de artistas plásticos del Colegio de Ciencias y Humanidades, El Rector de la UNAM, Doctor Jorge Carpizo, instala la Comisión que diagnosticará los principales problemas académicos y administrativos de la Unidad Académica del Ciclo de Bachillerato del Colegio de Ciencias y Humanidades.

1988

El Rector de la UNAM, Jorge Carpizo, designa al ingeniero Alfonso López Tapia, Coordinador del Colegio de Ciencias y Humanidades, se presenta el Plan de Trabajo 1988, con el fin de impulsar el desarrollo del Colegio, se suscribe un convenio don el Centro de Investigación y Servicio Educativo (CIDE), mediante el cual se formaliza el Programa de Formación para el Ejercicio de la Docencia en el Bachillerato del CCH.

Para los años 1989 y hasta la fecha no se logró obtener información, sin embargo, este trabajo fue inspirado, por el momento que vive el Colegio, en lo que se refiere a el cambio en los programas de estudio.

3.- ESTADISTICA

La estadística está considerada una ciencia, y su definición puede ser tan simple o tan complicada como simple o complicado sea el estudio que se esté efectuando, por ejemplo, si se desea saber ¿cuál es el número de vehiculos que pasan por una cierta caseta de cobro o el número de personas que escuchan una cierta estación de radio?. Se puede decir que:

Estadística.- *es la ciencia cuyo objeto es reunir, clasificar y contar todos los hechos en un mismo orden.*

Si por otro lado se quiere hacer un estudio sobre los índices de contaminación de la Ciudad de México, para proporcionar algunas alternativas de solución, es claro que también se está proyectando un estudio de tipo estadístico, pero ya no tan sencillo pues intervendrán en él muchas variables que lo complican. Y en este caso se puede definir como:

Estadística.- *Trata de la teoria y aplicación de métodos para coleccionar datos, analizarlos y hacer deducciones a partir de ellos.*

Otra definición posible sería:

Estadística.- *está ligada con los métodos científicos en la toma, organización, recopilación, presentación y análisis de datos, tanto para la deducción de conclusiones como para tomar decisiones razonables de acuerdo con tal análisis*

3.1 CONOCIMIENTO DE LOS FENÓMENOS ALEATORIOS

El estudio de los fenómenos deterministas y de los fenómenos aleatorios, sólo se *definen en este apartado y se profundizarán estos conceptos cuando se vea el tema Probabilidad.*

3.1.1 FENÓMENOS DETERMINISTAS Y FONÓMENOS ALEATORIOS

Fenómenos Deterministas.- *Son aquellos en los cuales de antemano se sabe cual será el resultado.*

EJEMPLO 3.1

I) Después del dia ¿qué viene?, pues la noche y a nadie se le ocurrirá pensar en otra cosa.

II) Si se junta un cerillo encendido y una sustancia inflamable ¿qué sucederá? Obviamente se dará el fuego y nadie pensaria que otra cosa pudiera suceder, por lo que se considera un hecho consumado.

Fenómenos Aleatorios.- *Para definir este tipo de fenómenos, es necesario entender que, hablar de aleatorio es lo mismo que hablar de azar o probabilidad, es decir, de incertidumbre, por lo tanto, éstos al contrario de los deterministas no se sabe que es lo que va a ocurrir.*

EJEMPLO 3.2

i) Cuando una persona se propone participar en una carrera de 20 kilómetros, hay dos posibles resultados de interés los cuales son, que termine la carrera y que no la termine, el punto está, en que no se sabe que resultado sucederá.

ii) Cuando un espermatozoide fecunda a un óvulo, hay dos posibles resultados de *interés, que sea mujer u hombre, pero hasta ahora no se saber que realmente sucederá. Posiblemente en futuro con el avance de la ciencia tal fenomeno pueda controlarse y conocerse.*

Cabe mencionar que antiguamente el parto, se consideraba un fenómenos aleatorio, pero hoy en día debido a las técnicas modernas de ultrasonido, se puede saber si el proceso de gestación materna corresponde al de una dama o no, por lo que el parto en si ha dejado de ser considerado un fenómeno aleatorio

3.1.2 ESTUDIO, SIMULACIÓN Y GENERACIÓN DE FENÓMENOS ALEATORIOS

Se han definido los fenómenos aleatorios y deterministas, pero ahora se empezará el estudio de la Estadística y el tratamiento de los fenómenos aleatorios, como se mencionó con anterioridad, serán estudiados en el capítulo de Probabilidad que son los que a ella le preocupan o interesan.

El hombre desde tiempo inmemorial, registra la información de diversos sucesos, con el propósito de estudiarlos y saber, ¿qué sucede?, ¿cómo sucede?, ¿cómo lo afecta?, etc. Tal es el caso de los nacimientos, las muertes, los asesinatos, las cosechas, las votaciones y un sin número de ejemplos más.

Por ese lado, la estadística ha hecho mucho al respecto y tiene más por hacer, sin embargo, hay ciertos fenómenos cuya ocurrencia o no es frecuente o bien puede ser catastrófica, como por ejemplo la quiebra de una empresa, la muerte de un paciente de una cierta enfermedad, entre otro. En dichos casos, el conocimiento humano ha desarrollado opciones alternativas para que a pesar de todo se pueda, en la medida de lo posible, conocer el comportamiento de dichos fenómenos bajo ciertas circunstancias, tal es el caso de la simulación por computadora, que consiste en desarrollar algoritmos que una vez programados semejan el fenómeno en estudio, naturalmente mientras mejor conceptualizado esté el fenómeno, seguramente los resultados obtenidos serán *más apropiados.*

La Estadística para su estudio se divide en:

para datos agrupados

descriptiva

para datos no agrupados

Estadística

inferencial

3.2 ESTADISTICA DESCRIPTIVA

La Estadística Descriptiva*, está relacionada con los métodos de recopilación, organización, presentación, análisis e interpretación de un grupo de datos, ya sea con el total de los datos o una parte de ellos llamada muestra.*

Se comenzará por estudiar la Estadística Descriptiva y posteriormente la Inferencial. Por tal motivo, es conveniente definir algunos conceptos que son indispensables para lograr tal propósito

3.2.1 VARIABLE, POBLACIÓN Y MUESTRA

VARIABLE .- Una variable desde el punto de vista estadístico es una característica que está "presente" o tienen todos los elementos de una poblacion y que varía o cambia de un elemento a otro. Ejemplos de lo anterior son lo siguientes

El sexo. Hay elementos que son o tienen la "característica" de mujer como también los hay que son hombres, esta caracteristica no tiene muchos "valores" o "categorías" sólo dos, pero cambia de un elemento a otro.

La ocupación, aquí los valores o categorias considerados pueden ser varios, por ejemplo albañil, carpintero, jardinero, empleado, profesor, gerente, secretaria, ama de casa, etcétera. Esta característica tiene más valores o categorías y suele determinarse o especificarse según el propósito de estudio en donde se vaya a utilizar dicha característica, aunque tenga o pueda tener varios valores no se consideran demasiados

El peso es otra característica que cambia de un elemento a otro de la población y los valores que pueden tomar son demasiados, casos de ellos pueden ser, 1350, 2150,... etcétera, representados en gramos, aunque se podrían reportar en otros múltiplos o submúltiplos , todo ello dependiendo de la población de que se trate y los propósitos del estudio.

Es importante hacer notar que siempre en un valor cualquiera, se estarán ubicando no sólo elementos de la misma población muy cercanos al valor ya considerado, esto es, en el caso del peso, habrá algunos cuyo valor será 57 kilogramos exactamente, pero seguramente habrá más cuyos valores oscilarán entre ± 0.000001 *ó* ± 0.00001 *ó* ± 0.0001 *ó* ± 0.001 *ó* ± 0.001 *ó* ± 0.01 *o cualquier combinación de los anteriores utilizando todos los dígitos que se conocen (1,2,3,4,5,6,7,8,9,0).*

La edad es otra característica que también tiene demasiados valores y que generalmente se suelen registrar, reportar o considerar de manera discreta, difícilmente a una persona se le ocurrirá decir que tiene 21 años, 3 meses, 24 días, 5 horas, 32 minutos y 51 segundos, cuando se le pregunte su edad, lo que muy a menudo ocurre es que se reportan los años y los meses o simplemente los años y en el supuesto caso de que a alguien se le ocurra contestar como se señaló anteriormente, por lo general lo que se suele registrar es lo ya comentado.

De lo anterior se pueden observar dos cosas sumamente importantes, la primera es que hay características que tienen como "valores" pertenecientes a lo intrínseco de su naturaleza ideas que no son estrictamente hablando números, como fue el caso de las características, sexo y ocupación, pero que también hay características que utilizan números para manifestarse o entenderse como fue el caso del peso y la edad.

EJEMPLO 3.3

i) El número de hijos varones en una cierta familia, puede tomar los valores de 0,1,2,... pero no puede tomar valores de 1.5 o 4.56 .

ii) El número de tornillos defectuosos de una producción de 1000 en una cierta empresa, en este caso como en el anterior, la variable puede tomar valores de 0,1,2,3...... .

iii) El número de focos de 60w. Producidos diariamente en México durante un mes.

iv) La temperatura registrada en el observatorio astronómico de México cada hora, durante un día.

v) El periodo de duración de los focos de 60 w., producidos por una empresa.

Una variable para la que sus posibles valores son únicamente categorías se llama Variable Categórica. por ejemplo, al evaluar la aceptacion de un cierto producto comestible de nueva creación, se les preguntó a las amas de casa si *consideraban que el nuevo producto era bueno, malo o regular.*

Las Variables Categóricas que no establecen ninguna relación de orden entre sus categorías, es llamada Variable Categórica Nominal.

Si una Variable Categórica establece alguna relación de orden entre sus categorías, es llamada Variable Categórica Ordinal

Si para una Variable, los posibles valores que puede tomar son números, se le llamará Variable Numérica.

Una Variable Numérica que se caracteriza por la propiedad de que para dos posibles valores de la variable solamente hay un número finito valores, se llama Variable Numérica Discreta.

Si una variable tiene la propiedad de que entre dos posibles valores de ésta cualquier valor intermedio es también un valor posible de la variable, entonces es llamada Variable Numérica Continua.

Es muy importante comprender el significado de variable, sin embargo, en estadística, es igualmente importante introducir el concepto de variable aleatoria que será definida perfectamente bien cuando se vea Probabilidad.

POBLACIÓN .- Es el conjunto de todos los posibles elementos. La población es un conjunto, lo que indica que debe ser bien definido y cuya característica primordial es que es universal.

La población puede ser finita o infinita. Finita cuando sus elementos pueden ser contables y numerables e infinita cuando sus elementos pueden ser numerables pero no contables. Algunos estudios estadísticos son realizados con base en la población, se dice que se esta realizando un censo.

EJEMPLO 3.4

- *El grupo 812 de estadística I de la Escuela Nacional Colegio de Ciencias y Humanidades plantel Vallejo. (Población finita).*
- *Los microorganismos que se pueden encontrar en una sustancia contaminada en un cierto laboratorio. (Población infinita)*

MUESTRA - Es un subconjunto de la población.

Cuando un ensayo se lleva a cabo con los datos de una muestra, se dice que se esta muestreando, o realizando un estudio por muestreo.

Obviamente, la mayoría de los trabajos se efectúan con muestreo ya que éste resulta menos costoso y reduce el margen de error.

EJEMPLO 3.5

i) Diez alumnos tomados al azar del grupo 812 de Estadística I de la Escuela Nacional Colegio de Ciencias y Humanidades plantel Vallejo.

ii) 35,000 habitantes de la Ciudad de México.

iii) 1000 tornillos producidos por una máquina en una cierta empresa.

3.2.2 RECOPILACIÓN Y ORGANIZACIÓN DE DATOS

Las dos vertientes por las cuales se pueden recopilar, organizar y analizar datos en Estadística Descriptiva son, cuando los datos se toman en grupo (datos agrupados) y cuando se toman unitariamente (datos no agrupados). Tanto en el primer caso como en el segundo, se debe tener cuidado, al seleccionar los datos, que éstos sean significativos, es decir, que puedan proporcionar buenos resultados no tendenciosos

Por ejemplo si se quisiera hacer un estudio sobre el desempleo en México, y se seleccionaran sólo personas que tienen empleo, es muy probable que los datos *recopilados no dieran los resultados confiables.*

Tomando en primera instancia cuando se tratan datos no agrupados, la forma de recopilarlos es casi siempre sencilla, pues muchos de ellos salen de la observación repetida de u suceso, por ejemplo, las calificaciones que una persona obtiene en un semestre, el número de goles que un jugador de fútbol pueda anotar en diez partidos jugados, las veces que llueve en el año, etcétera.

EJEMPLO 3.6

Un equipo de fútbol profesional, contrató a prueba a un goleador, se acordó que se observaria su comportamiento en diez partidos jugados, obteniéndose los siguientes datos.

JUEGO No.	*GOLES ANOTADOS*
1	*1*
2	*0*
3	*0*
4	*1*
5	*1*
6	*0*
7	*0*
8	*0*
9	*0*
10	*0*

ya organizados los datos en forma tabular, se podrá hacer un pequeño análisis que aunque elemental ya se puede considerar estadístico. Al observar la tabla anterior, fácilmente se puede ver que el jugador sólo anotó 3 goles en diez partidos, por lo que se puede tomar una decisión.

Otro ejemplo muy sencillo que puede mostrar la recopilación y la organización de datos es el siguiente:

EJEMPLO 3.7

En la clase de Estadística y Probabilidad I, el profesor llevó a cabo cinco exámenes para evaluar el curso, él escogió al azar a uno de sus alumnos y observó las calificaciones que obtuvo en sus cinco exámenes

No. de examen	*Calificación*
1	*8*
2	*5*
3	*10*
4	*6*
5	*5*

Como fácilmente se puede observar, los datos que arroja este pequeño estudio son pocos, pero suficientes para obtener el promedio de el alumno escogido.

EJEMPLO 3.8

Un obrero que trabaja por su cuenta, quiso llevar un control sobre sus entradas durante una semana y elaboró la siguiente tabla.

Días	***entradas económicas***
lunes	*426.00*
martes	*618.50*
miércoles	*298.00*
jueves	*98.00*
viernes	*547.00*

los datos anteriores muestran que el obrero ningún día ganó la misma cantidad, lo que significa que los datos aparecen sólo una vez , con lo cual la recopilación y organización es elemental y sencilla. Sin embargo, los datos estadísticos en su mayoría son encontrados en mayor cantidad, dificultando aun más las formas de organización y de recopilación. Con los siguientes ejemplos se tratará de mostrar otras formas de organización para cuando los son muchos.

EJEMPLO 3.9

Se llevó a cabo un estudio con 20 personas fumadoras que están en un proceso de desintoxicación de un año y se requiere saber después de seis meses ¿cuál ha sido el avance obtenido? Si antes del tratamiento, cada uno de ellos fumaba un promedio de 40 cigarrillos (dos cajetillas de cigarros) diarios.

La forma como se obtuvieron los datos, fue preguntando a cada uno de ellos ¿cuantos cigarros fuma en este momento? De igual forma, una manera de organizarlos, es tomando sus nombres

Nombre del Fumador	*Número de cigarros consumidos por dia*
Arturo Canales R.	*4*
Juan Escalante	*8*
Maria Pérez	*6*
Josefina Ron	*4*
Roberto Tindero	*6*
Ernesto Lecuona	*6*
Rubén Escalona	*4*
Alicia Rubio	*8*
Juan Matin P.	*6*
Martha Mora	*7*
Alberto Martínez	*3*
Estela Vargas	*6*
Teresa Barrón	*3*
Carlos Sandoval	*3*
Alberto Gómez	*7*
Raquel Olmos	*7*
Ramon Escalante O	*8*
Hilda Castro	*8*
Carmen Rosas	*4*
Ramiro Cabrera	*4*

Si se observa la tabla anterior con detenimiento, se puede ver que varias personas fuman a la fecha 3 cigarros, lo mismo sucede con aquellos que fuman 4 o más.

En muchas ocasiones para poder organizar y analizar mejor los datos, pueden ser agrupados por alguna característica de importancia y que es común en el grupo. Por ejemplo, se podrían agrupar por número de cigarros fumados al día, es decir se pueden formar grupos de personas que fumen la misma cantidad de cigarros por día quedando de la siguiente manera.

Cigarros consumidos por día.	*Conteo*	*Número de personas que pertenecen al grupo*
3	*xxx*	*3*
4	*xxxxx*	*5*
5		*0*
6	*xxxxx*	*5*
7	*xxx*	*3*
8	*xxxx*	*4*

El conteo, como su nombre lo indica, consiste en contar el número de personas que pertenecen a cada grupo, lo que es representado con una cruz. La suma de las cruces o número de personas pertenecientes a cada grupo, será llamada ***frecuencia****.*

Analizando los datos de la tabla anterior, tres personas han logrado disminuir a

3 cigarros diarios, cuatro personas fuman ahora 5 cigarrillos. En el caso de los que fuman cinco, nadie fuma esta cantidad. 6 fuman cinco personas, 7 tres y 8 solo cuatro de los encuestados.

Existen dos formas de organizar datos agrupados, cuando los datos se pueden organizar en observaciones puntuales, y cuando estos por necesidad deben ser organizados por intervalos, llamados "intervalos de clase". Las formas en como estos pueden ser formados son varias, una de ellas se mostrará con el siguiente ejemplo

EJEMPLO 3.10

Se tomó una muestra de 50 personas de una escuela preparatoria y se midio su estatura en metros

1.62	1.54	1.91	1.70	1.70
1.80	1.68	1.81	1.85	1.78
1.92	1.69	1.72	1.59	1.50
1.56	1.68	1.70	1.64	1.87
1.58	1.79	1.75	1.73	1.80
1.49	1.49	1.80	1.73	1.60
1.82	1.87	1.53	1.80	1.65
1.63	1.58	1.59	1.62	1.71
1.66	1.69	1.60	1.67	1.59
1.73	1.79	1.63	1.50	1.80

Siguiendo un metodo de organización por pasos en primer punto, se obtendrá el **Rango,** *que consiste en la diferencia que existe entre el dato mayor y el dato menor.*

$$Rango = dato\ mayor - dato\ menor$$

$$= 1.92 - 1.49$$

$$= .43$$

Esta medida al ser dividida entre el número de intervalos que se deseen formar, proporciona sólo un criterio de apertura de intervalos, lo que significa que, la mayoría de las veces no es exacto, por lo que se tendrá que ajustar. Supóngase que se quieren formar cinco intervalos.

$$Apertura\ del\ intervalo = \frac{Rango}{número\ de\ intervalos}$$

$$= \frac{.43}{5}$$

$$= .09$$

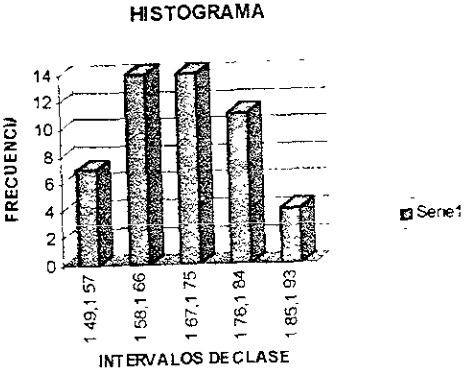
es decir, los intervalos deben abrirse en nueve centésimas. En ocasiones esta medida es exacta y al formar los intervalos, los datos son contenidos totalmente en ellos, sin embargo, en la mayoría de las veces, esto no sucede por lo que se puede tomar la decisión de abrir en menos .08 o en mas grande .10, en este caso se abrirán a .09 mts., se formara también el conteo (datos que pertenecen a cada grupo) y la frecuencia fi. Se debe aclarar que la abertura es puntual.

Intervalo de clase	conteo	frecuencia fi
1 49 - 1.57	xxxxxxx	7
1.57 - 1.66	xxxxxxxxxxxxxx	14
1.67 - 1.75	xxxxxxxxxxxxxx	14
1.76 - 1.84	xxxxxxxxxxx	11
1.85 - 1 93	xxxx	4

3.2.3 TABLAS DE DISTRIBUCIÓN DE FRECUENCIA Y REPRESENTACIÓN GRÁFICA.

Una ves recopilados y organizados los datos, el principal interes consiste en observar cual es su comportamiento, una de varias formas, es por medio de graficas, las más comunes son.

- ***HISTOGRAMA O HISTOGRAMA DE FRECUENCIA.-*** *Consiste en una serie de rectángulos, los cuales tienen sus bases igual al tamaño de los intervalos de clase y altura igual a las frecuencias. Continuando con el ejemplo, se tiene*



Se puede obtener del cuadro anterior las frecuencias acumuladas fa que son calculadas de la siguiente manera

$$fa_1 = f_1$$

$$fa_2 = f_1 + f_2$$

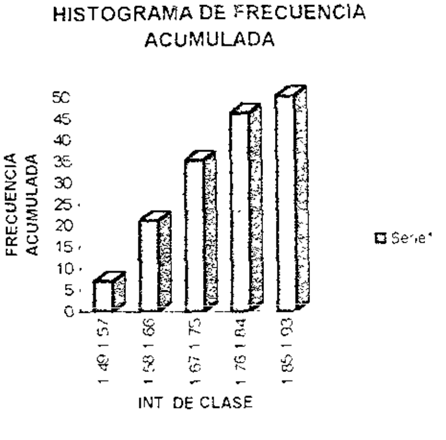
$$fa_3 = f_1 + f_2 + f_3$$

$$fa_4 = f_1 + f_2 + f_3 + f_4$$

$$fa_5 = f_1 + f_2 + f_3 + f_4 + f_5$$

Intervalo de clase	*frecuencia fi*	*frec. Acum. fa*
1.49 - 1.57	*7*	*7*
1.57 - 1.66	*14*	*21*
1.67 - 1.75	*14*	*35*
1.76 - 1.84	*11*	*46*
1.85 - 1.93	*4*	*50*

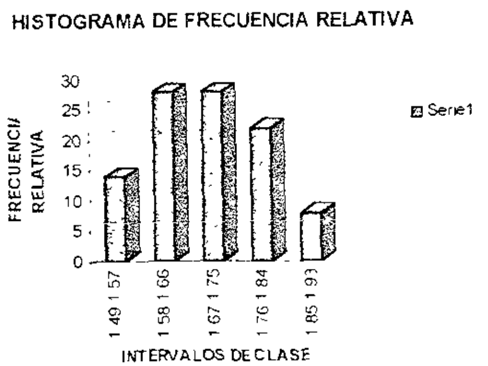
Con las cuales se puede obtener el histograma de frecuencia acumulada



Así como fueron obtenidas la frecuencias acumuladas se puede llevar a cabo otro grafico dentro de la familia de los histogramas y este es el histograma de frecuencia relativa fr. la cual consiste en los porcentajes que tienen las frecuencias fi con respecto a la suma de las frecuencias

Intervalo de clase	*frecuencia fi*	*frec. Rel. fr*	
1.49 - 1.57	*7*	*7 50*	*.14*
1.57 - 1.66	*14*	*14 50*	*.28*
1.67 - 1.75	*14*	*14 50*	*.28*
1.76 - 1.84	*11*	*11 50*	*11*
1.85 - 1.93	*4*	*4 50*	*.08*

de donde.



- ***POLÍGONO DE FRECUENCIA.-*** *Es un grafico formado por una serie de puntos obtenidos entre las marcas de clase Xi y la frecuencias fi, unidos estos por una linea recta entre si*

Para obtener esta gráfica, se deben encontrar las ***marcas de clase Xi***. *Es conveniente aclarar que todo intervalo de clase consta de dos límites el de el lado izquierdo llamado "límite inferior de clase" y el de el lado derecho llamado "límite superior de clase". La marca de clase será entonces el resultado de la suma de los límites inferior de clase y superior de clase divididas entre dos.*

$$X_1 = \frac{149 + 157}{2} = 153$$

$$X_2 = \frac{158 + 166}{2} = 162$$

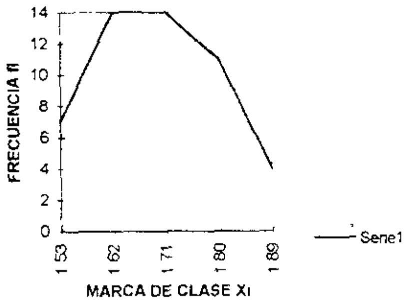
$$X_3 = \frac{167 + 175}{2} = 171$$

$$X_4 = \frac{176 + 184}{2} = 180$$

$$X_5 = \frac{185 + 193}{2} = 189$$

Intervalo de clase	*frecuencia fi*	*m.c. Xi*
1.49 - 1.57	*7*	*1.53*
1.57 - 1.66	*14*	*1.62*
1.67 - 1.75	*14*	*1.71*
1.76 - 1.84	*11*	*1.80*
1.85 - 1.93	*4*	*1.89*

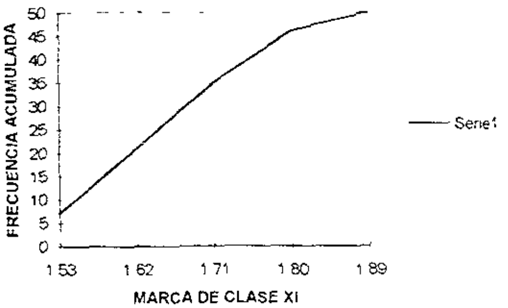
POLÍGONO DE FRECUENCIA



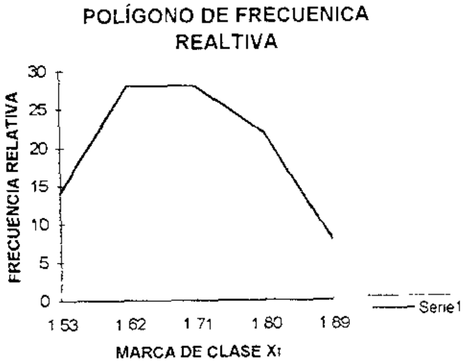
Otro tipo de polígono, es el que se forma siguiendo el mismo método que el anterior pero tomando en cuenta las frecuencias acumuladas, de ahí su nombre, ***polígono de frecuencia acumulada.***

Xi	*fai*
1 53	*7*
1.62	*21*
1.71	*35*
1 80	*46*
1 89	*50*

POLÍGONO DE FRECUENCIA ACUMULADA



Así mismo, con las frecuencias relativas se formara el ***polígono de frecuencia relativa.***



En muchos casos reales, las representaciones gráficas proporcionan un punto de vista sencillo del comportamiento de los datos. Así un grafico que es muy comun de uso muy cotidiano y de muy buen aspecto, es la gráfica circular o grafica de pastel. La forma en como se construye este tipo de graficas es:

Se lleva a cabo una equivalencia proporcional de las frecuencias con la medida en grados de una circunferencia, lo que se logra, por medio de una regla de tres simple. Ejemplo.

Para el primer caso

Frecuencia	*grados*
50	*360*
7	*x*

de donde:

$$x = \frac{360 * 7}{50} = 50.4 \approx 50°$$

para el segundo caso

$$\begin{array}{cc} 360 & 50 \\ x & 14 \end{array}$$

$$x = \frac{360 * 14}{50} = \frac{5040}{50} = 100\ 80 \approx 101^0$$

para el tercer intervalo

$$\begin{array}{cc} 360 & 50 \\ x & 14 \end{array}$$

$$x = \frac{360 * 14}{50} = \frac{5040}{50} = 100\ 80 \approx 101^0$$

para el cuarto

$$\begin{array}{cc} 360 & 50 \\ x & 11 \end{array}$$

$$x = \frac{360 * 11}{50} = \frac{3960}{50} = 79\ 20 \approx 79$$

por último

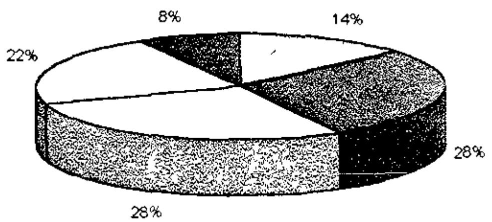
$$\begin{array}{cc} 360 & 50 \\ x & 4 \end{array}$$

$$x = \frac{360 * 4}{50} = \frac{1140}{50} = 28\ 80 \approx 29$$

Intervalo de clase	*frecuencia fi*	*grados*
1.49 - 1.57	*7*	*50*
1.57 - 1.66	*14*	*101*
1.67 - 1.75	*14*	*101*
1.76 - 1.84	*11*	*79*
1.85 - 1.93	*4*	*29*

graficando

GRÁFICA CIRCULAR O DE PASTEL

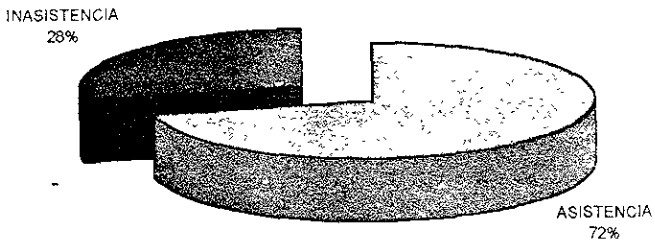


Otro ejemplo del uso de la grafica de pastel, lo proporciona el porcentaje de asistencia de los profesores en el Colegio de Ciencias y Humanidades Plantel Vallejo tomado en febrero de 1998.

Asistencia 72%

Inasistencia 28%

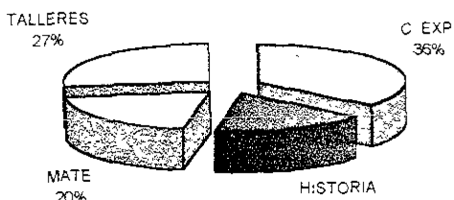
PORCENTAJE DE ASISTENCIA DEL COLEGIO DE CIENCIAS Y HUMANIDADES PLANTEL VALLEJO FEBRERO DE 1998



De igual forma, el porcentaje de inasistencia por área fue de:

Ciencias Experimentales	*36%*
Historia	*17%*
Matemáticas	*20%*
Talleres	*27%*

PORCENTAJE DE INASISTENCIA POR ÁREA FEBRERO DE 1998



A continuacion se efectuara un ejemplo tratando de que con este se puedan madurar el conocimiento de lo estudiado

EJEMPLO 3.11

Los siguientes datos, representan una muestra en grados centigrados de una cierta materia

47.20	46.54	46.42	46.60	46.89	47.40	46.90	46.30
47.40	46.90	47.12	46.88	46.87	46.53	46.93	46.80
47.10	46.69	46.90	47.22	46.96	46.48	47.00	46.54
46.94	46.35	46.87	46.78	46.81	46.90	46.52	46.89

Se obtendrá:

1 El rango

2 Cinco grupos o intervalos de clase

3 Conteo (distribucion de los datos por grupo)

4 Frecuencia fi (numero de datos por intervalo de clase)

5 Frecuencia Acumulada fa

6 Frecuencia Relativa fr

7 Histograma

8 Poligono

9 Grafica de Pastel

*1 **Rango***

Como el rango de defino como la diferencia del dato mayor y el dato menor, entonces

$$rango \quad 47.40 - 46.30$$

$$1.10$$

*2 **Intervalos de clase.***

*3 **Conteo.***

*4 **Frecuencia.***

Para obtener los intervalos de clase, el rango proporciona un excelente criterio de apertura, al dividirlo entre el número de intervalos deseados

$$\text{Tamaño del intervalo} \quad \frac{\text{rango}}{\text{numero de intervalos}}$$

$$\frac{1.10}{5}$$

$$.22$$

Es decir, los intervalos podrían ser abiertos en .22 , sin embargo en este ejemplo, al abrirse en esta medida los datos no son totalmente aceptados, por lo que se tendrá que tomar la decisión de dar una apertura de .23 quedando los intervalos de la siguiente manera. Ademas en el siguiente cuadro se muestra el conteo y la frecuencia fi

Intervalos de clase	*conteo*	*frecuencia fi*
46.30 - 46.52	xxxx	4
46.53 - 46.75	xxxx	4
46.76 - 46.98	xxxxxxxxxxxxxxxx	16
46.99 - 47.21	xxxx	4
47.22 - 47.44	xx	2

5 ***Frecuencia Acumulada fa.***

La frecuencia acumulada como ya se mencionó, consiste en acumular las frecuencias por intervalo

Intervalos de clase	*frecuencia fi*	*frec. Acum. fa*
46.30 - 46.52	4	4
46.53 - 46.75	4	8
46.76 - 46.98	16	24
46.99 - 47.21	4	28
47.22 - 47.44	2	30

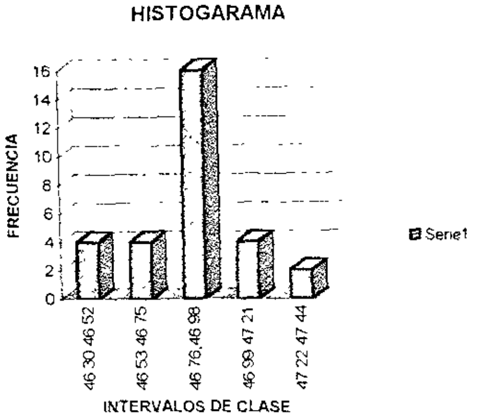
6 ***Frecuencia relativa fr.***

Hablar de frecuencia relativa es hablar de porcentajes, de tal manera que la suma de las frecuencias relativas tiene que ser igual a 1

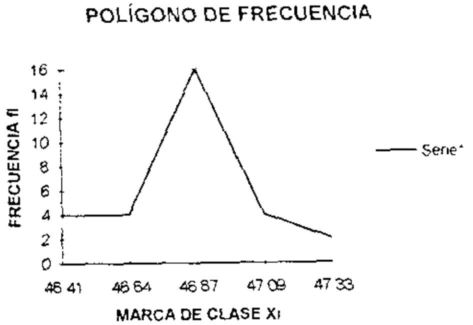
Intervalos de clase	frecuencia fi	frec. Rel. Fr.	
46 30 - 46.52	4	4 30	13
46.53 - 46.75	4	4 30	13
46 76 - 46.98	16	16 30	.53
46 99 - 47.21	4	4 30	.13
47.22 - 47 44	2	2 30	06

46 99

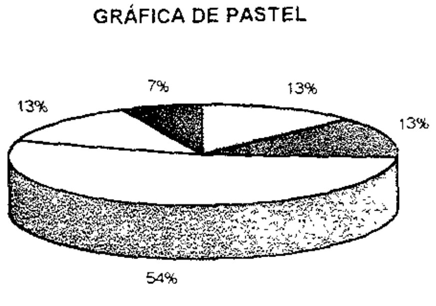
7. ***Histograma.***



8. *Polígono de frecuencia.*



9. *Gráfica de Pastel.*



3.2.5 *ASIGNACIÓN DE VALORES CARACTERÍSTICOS.*

Medidas de tendencia central

Las medidas de tendencia central, son parametros estadisticos que permiten observar entre otras cosas el bueno o mal comportamiento de los datos. Por ejemplo, supongase que en una poblacion determinada de México, el promedio (media aritmetica) por persona de gripe es una por año. Una persona cualquiera que pertenezca a esa poblacion seria normal que se enfermera de gripe una vez al año o una vez cada dos años o quizas una cada tres años. Pero no seria tan normal que una persona que pertenezca a esa poblacion se enfermara de gripe 6 o 7 veces al año o por otro lado que no se enfermara nunca si esta conviviendo con el grupo.

Como se puede observar, este pequeño analisis se pudo llevar a cabo gracias a que se contaba con el promedio de enfermos de gripe por año, que a continuacion de vera que es una medida de tendencia central.

MEDIA ARITMETICA.- La media aritmética es el tipo mas comun de los promedios que se pueden obtener en estadistica descriptiva y se define como:

Para datos no agrupados

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Donde:

$\bar{X}$ = media aritmetica

X_i = dato iesimo

$\sum_{i=1}^{n} X_i$ = la suma de todos los datos desde el primero hasta el enesimo

n = numero de datos

Para datos agrupados

$$\bar{X} = \frac{f_1 X_1 + f_2 X_2 + \dots + f_r X_r}{f_1 + f_2 + \dots + f_r} = \frac{\sum_{i=1}^{n} f_i X_i}{\sum_{i=1}^{n} f_i}$$

Donde

$\bar{X}$	*media aritmética*
Xi	*marca de clase*
fi	*frecuencia iésima*
$\sum_{i=1}^{n} f_i X_i$	*la suma de todas las multiplicaciones de las frecuencias por las marcas de clase Xi*
$\sum_{i=1}^{n} f_i$	*la suma de todas las frecuencias*

Primero se verá el caso de datos no agrupados tanto para la media aritmética mediana y moda, y posteriormente para datos agrupados

EJEMPLO 3.12

Retomando el ejemplo del futbolista tratado en el apartado 3 2 2, se tienen los siguientes datos

Un equipo de futbol profesional, contrato a prueba a un goleador, se acordó que se observaría su comportamiento en diez partidos jugados, obteniéndose los siguientes datos

JUEGO No	GOLES ANOTADOS
1	1
2	0
3	0
4	1
5	1
6	0
7	0
8	0
9	0
10	0

Como facilmente se puede observar, los datos que se presentan estan en forma unitaria, es decir no agrupados o se presentan uno por uno, lo que significa que en este caso para poder obtener la media aritmetica, se utilizara la formula para cuando los datos no estan agrupados.

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\bar{X} = \frac{1 + 0 + 0 + 0 + 1 + 1 + 0 + 0 + 0 + 0}{10}$$

$$\overline{X} = \frac{3}{10} = 30$$

EJEMPLO 3.13

Un estudiante del Colegio de Ciencias y Humanidades, en la materia de Estadística y Probabilidad I presentó cinco exámenes durante el semestre, obteniendo las siguientes calificaciones, en el primer examen su calificación fue de nueve, en el segundo de 8 3, en el tercero 7 8, en el cuarto de 10 y en el quinto y último 9 ¿Cual es su promedio o media aritmética?

Acomodando los datos se tiene:

Xi	*dato*
X1	*9.0*
X2	*8.3*
X3	*7.8*
X4	*10.0*
X5	*9.0*

calculando el promedio o media aritmetica.

$$\overline{X} = \frac{X_1 + X_2 + \quad + X_n}{n} = \frac{\sum_{i=1}^{n} Xi}{n}$$

$$\overline{X} = \frac{9\,0 + 8\,3 + 7\,8 + 10\,0 + 9\,0}{5}$$

$$\overline{X} = \frac{44\,10}{5} = 8\,82$$

EJEMPLO 3.14

Los Kilómetros recorridos por cinco estudiantes para venir al Colegio de Ciencias y Humanidades, Plantel Vallejo desde su casa, se muestran en el siguiente cuadro.

Estudiante	*kms. Recorridos (variable Xi)*
1	7
2	4
3	15
4	22
5	2

como el número de datos es impar, se utilizará la fórmula que corresponde, para lo cual, se ordenan los datos de menor a mayor.

$$2,\quad 4,\quad 7,\quad 15,\quad 22$$
$$X_1,\ X_2,\ X_3,\ X_4,\ X_5$$

sustituyendo

$$\widetilde{X} = X_{\left(\frac{n+1}{2}\right)}$$

$$\widetilde{X} = X_{\left(\frac{5+1}{2}\right)} = X_{\left(\frac{6}{2}\right)} = X_3 = 7$$

EJEMPLO 3.15

Se tienen los siguientes datos, 34, 23, 56, 89, 23, 95, y se quiere calcular la *mediana Ordenándolos de menor a mayor*

$X1$	$X2$	$X3$	$X4$	$X5$	$X6$
23	23	34	56	89	95

como el numero de datos es par, se tiene.

$$n = 6$$

$$X_{(\frac{n}{2})} = X_{(\frac{6}{2})} = X_3 = 34$$

$$X_{(\frac{n}{2}+1)} = X_{(\frac{6}{2}+1)} = X_{(3+1)} = X_4 = 56$$

como la fórmula para obtener la mediana para cuando el número de datos es par es

$$\tilde{X} = \frac{X_{(n/2)} + X_{(n/2+1)}}{2}$$

sustituyendo

$$\tilde{X} = \frac{34 + 56}{2} = \frac{90}{2} = 45$$

MODA.- La moda se conoce como el dato que tiene mayor frecuencia es decir aquel que aparece mas veces La moda puede existir o no, puede ser unica o haber mas de una

EJEMPLO 3.16

Primer caso. *La moda no existe.*

Supónganse los siguientes datos

1, 9, 10, 6, 8, 3

por lo que se puede observar, ninguno de los datos se repite más de una vez, por lo tanto se considera que la moda en este caso, no existe Otra forma en la cual no hay moda, es cuando los datos se repiten varias veces pero de igual manera.

1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3 3, 4, 4, 4, 4, 10, 10, 10, 10

Segundo caso. *La moda existe y es unica*

Los siguientes datos, representan una muestra al azar de 11 alumnos de una escuela primaria a los que se les preguntó su edad

9, 5, 5, 5, 9, 6, 8, 5, 9, 5, 8

como 5 es el dato que mas veces se repite, el valor de la moda es 5 y seria unica.

Tercer caso. *Mas de una moda*

Supónganse los siguientes datos

2, 7, 7, 7, 2, 2, 5, 12, 7, 7, 2, 2

en este caso el numero 7 y el 2 son los que mas se repiten y ademas de igual manera lo que hace que existan dos modas, dando un caso bimodal

Nota Pueden encontrarse mas de dos modas en un conjunto de datos, lo que se conoce como multimodal

DATOS AGRUPADOS

Aun cuando ya se han mencionado algunas cosas sobre datos agrupados ahora se tocara el caso de la media aritmetica, mediana y moda para tal efecto

MEDIA ARITMETICA

La media aritmética para datos agrupados se define como

$$\bar{X} = \frac{\sum_{1}^{r} f_i X_i}{\sum_{1}^{r} f_i}$$

donde

$\bar{X}$ *media aritmetica*

$\sum_{1}^{r} f_i X_i$ *suma de todas las multiplicaciones entre la marca de clase X_i y las frecuencias f_i*

$\sum_{1}^{r} f_i$ *suma de todas las frecuencias f_i*

EJEMPLO 3.17

La siguiente tabla, muestra la forma en como fueron recopilados y organizados 30 datos en grados centigrados

Intervalos de clase	*frecuencia fi*
46.30 - 46.52	4
46.53 - 46.75	4
46.76 - 46.98	16
46.99 - 47.21	4
47.22 - 47.44	2

limites inferiores de clase

limites superiores de clase

Se obtendrá, la media aritmética, mediana y moda

Media aritmética $\overline{X}$.

Para calcular la media aritmética, se deben calcular las marcas de clase que se obtienen de la suma de los limites inferiores mas los superiores divididos entre 2

Intervalos de clase	*fi*	*marca de clase Xi*
46.30 - 46.52	4	46.41
46.52 - 46.75	4	46.64
46.76 - 46.98	16	46.87
46.99 - 47.21	4	47.09
47.22 - 47.44	2	47.33

se definió la media aritmética como

$$\bar{X} = \frac{\sum_{i=1}^{5} f_i X_i}{\sum_{i=1}^{5} f_i}$$

desarrollando

$$\bar{X} = \frac{f_1X_1 + f_2X_2 + f_3X_3 + f_4X_4 + f_5X_5}{f_1 + f_2 + f_3 + f_4 + f_5}$$

$$\bar{X} = \frac{(4)(46.41) + (4)(46.64) + (16)(46.87) + (4)(47.09) + (2)(47.33)}{4 + 4 + 16 + 4 + 2}$$

$$\bar{X} = \frac{185.64 + 186.56 + 749.92 + 188.36 + 94.66}{30}$$

$$\bar{X} = \frac{1405.14}{30} = \mathbf{46.84}$$

MEDIANA

Por lo que se refiere a la mediana para datos agrupados, existen algunos criterios de medida, por una parte se puede llegar a conocer únicamente el intervalo de clase en donde se encuentra la mediana, o dar un valor aproximado a la mediana por medio de interpolación.

Para ejemplificar lo mencionado se tomaran los datos del ejemplo anterior, calculando ademas frecuencia acumulada e intervalos reales de clase que sirven para lograr tal efecto

EJEMPLO 3.18

Int. de clase	*fi*	*fa*	*Xi*	*Int. Reales de clase*
46.30 - 46.52	4	4	46.41	46.295 - 46.525
46.52 - 46.75	4	8	46.64	46.525 - 46.755
46.76 - 46.98	16	24	46.87	46.755 - 46.985
46.99 - 47.21	4	28	47.09	46.985 - 47.225
47.22 - 47.44	2	30	47.33	47.225 - 47.445

La formula para obtener la mediana es

$$\tilde{X} = L_i + \left(\frac{\frac{\sum_{i=1}^{n} fi}{2} - (\sum f)_i}{fmediana} \right) c$$

donde

$\tilde{X}$ = *mediana*

L_i = *limite real inferior del intervalo de clase que contiene a la mediana*

$\sum_{i=1}^{n} fi$ = *suma de todas las frecuencias fi*

$(\sum f)_1$ = *suma de todas las frecuencias por debajo de la frecuencia del intervalo de clase que contiene a la mediana*

f mediana = frecuencia del intervalo que contiene a la mediana

c = tamaño del intervalo que contiene a la mediana

Como antes se mencionó, se puede únicamente conocer el intervalo que contiene a la mediana, con lo que muchos quedan satisfechos Sin embargo, aplicando la fórmula anterior, se puede llegar por interpolación al valor de la mediana

Ya se dijo que la mediana es el valor que divide en dos partes iguales al conjunto de datos, por lo que para encontrar el intervalo que contiene a la mediana, se divide la suma de todas las frecuencias entre 2

$$\frac{\sum_{i=1}^{i=5} fi}{2} = \frac{30}{2} = 15$$

el resultado obtenido (15) se busca cual es la primera frecuencia acumulada de manera ascendente que lo contiene.

Int. de clase	*fi*	*fa*	*Xi*	*Int. Reales de clase*
46.30 - 46.52	*4*	*4*	*46.41*	*46.295 - 46.525*
46.52 - 46.75	*4*	*8*	*46.64*	*46.525 - 46.755*
46.76 - 46.98	*16*	*24*	*46.87*	*46.755 - 46 985*
46.99 - 47.21	*4*	*28*	*47.09*	*46.985 - 47 225*
47.22 - 47.44	*2*	*30*	*47.33*	*47.225 - 47.445*

frecuencia acumulada que contiene a 15

*En este caso, 24 es la frecuencia acumulada que contiene a 15, por lo cual, el intervalo de clase que contiene a la mediana es (46 76 - 46.98) y el intervalo real de clase que contiene a la mediana sería (46.755 - 46.985), quedando así que el límite inferior real de clase que contiene a la mediana es: **46.755***

Por lo que

$$L_l = 46.755$$

así mismo.

$$\frac{\sum_{i=1}^{i=5} fi}{2} = \frac{30}{2} = 15$$

la frecuencia del intervalo que contiene a la mediana es 16, la suma de las frecuencias que están por debajo de ésta, son: 4 + 4 = 8 por lo que.

$$(\sum f)_1 = 4 + 4 = 8$$

además

$$f\,mediana = 16$$

por último

$$c = .23$$

sustituyendo

$$\tilde{X} = 46\,755 + (\frac{15-8}{16})23$$

$$\tilde{X} = \mathbf{46.84}$$

MODA

La moda de un conjunto de datos, consiste en el dato que más veces se repite o tiene más frecuencia. Para calcular la moda, sucede algo parecido a lo que sucedió con la mediana, que puede ser que algunos se conformen con saber cual es intervalo de clase que la contiene, pero también en este caso se tratará de dar una opción para acercarse más al valor de la moda, aplicando la siguiente fórmula.

$$\textbf{Moda} = L_I + \frac{\Delta_1}{\Delta_1 + \Delta_2}\, c$$

donde:

L_I = *límite real inferior de la clase que contiene a la moda*

Δ_1 = *exceso de frecuencia, de la frecuencia del intervalo que contiene a la moda con respecto a la frecuencia inmediata inferior*

Δ_2 = *exceso de frecuencia, de la frecuencia del intervalo que contiene a la moda con respecto a la frecuencia inmediata superior*

c = *tamaño del intervalo de clase que contiene a la moda*

Siguiendo con los datos del ejemplo anterior.

EJEMPLO 3.19

En primera instancia, se debe encontrar el intervalo de clase que contiene a la *moda. Como la moda es el dato que más se repite, el intervalo de clase que contiene a la moda será aquel que tenga más frecuencia.*

Int. de clase	*fi*	*fa*	*Xi*	*Int. Reales de clase*
46.30 - 46.52	*4*	*4*	*46.41*	*46.295 - 46.525*
46.52 - 46.75	*4*	*8*	*46.64*	*46.525 - 46.755*
46.76 - 46.98	***16***	*24*	*46.87*	*46.755 - 46 985*
46.99 - 47.21	*4*	*28*	*47.09*	*46.985 - 47.225*
47.22 - 47.44	*2*	*30*	*47.33*	*47.225 - 47.445*

intervalo de clase que contiene a la moda

así, el límite real inferior que contiene a la moda es. 47.755 es decir

$$L_l = 46.755$$

para obtener Δ_1 *y* Δ_2 *los excesos de frecuencia, se hace lo siguiente, como 16 es la frecuencia del intervalo que contiene a la moda, que además es la mayor, ésta se debe restar a la frecuencia del intervalo inmediato inferior, que en este caso sería 4, de donde*

$$\Delta_1 = 16 - 4 = 12$$

Int. de clase	*fi*	*fa*	*Xi*	*Int. Reales de clase*
46.30 - 46.52	*4*	*4*	*46.41*	*46.295 - 46.525*
46.52 - 46.75	*4*	*8*	*46.64*	*46.525 - 46.755*
46.76 - 46.98	*16*	*24*	*46.87*	*46.755 - 46.985*
46.99 - 47.21	*4*	*28*	*47.09*	*46.985 - 47.225*
47.22 - 47.44	*2*	*30*	*47.33*	*47.225 - 47.445*

frecuencia del intervalo inmediato inferior

frecuencia del intervalo inmediato superior *por otro lado, al restarle la frecuencia del intervalo inmediato superior, queda*

$$\Delta_2 = 16 - 4 = 12$$

en este ejemplo, $\Delta_1 = \Delta_2$ *cosa que no sucede con frecuencia*

Por último

$$c = .23$$

sustituyendo en la fórmula.

$$\textbf{Moda} = 46.755 + \left(\frac{12}{12+12} \right) .23$$

$$= 46.755 - .115$$

$$= \mathbf{46.87}$$

CUARTILES

Los cuartiles son medidas estadísticas que dividen a un conjunto de datos ordenados en cuatro partes iguales. Aun cuando el conjunto de datos queda dividido en cuatro partes iguales como se mencionó, los cuartiles son tres, Q_1, Q_2 *y* Q_3 *, y se calculan de la siguiente manera·*

Primer cuartil Q_1.

$$Q_1 = L_{Q_1} + \left(\frac{\dfrac{\sum_{i=1}^{i=n} fi}{4} - (\sum f)_{Q_1}}{f_{Q_1}} \right) c$$

donde

Q_1 = *primer cuartil*

L_1 = *límite real inferior del intervalo que contiene al primer cuartil.*

$\sum_{i=1}^{i=n} fi$ = *suma de todas las frecuencias*

$\left(\sum f\right)_{Q_1}$ = *suma de todas las frecuencias por debajo de la frecuencia del intervalo que contiene al primer cuartil.*

c = *tamaño del intervalo que contiene al primer cuartil.*

EJEMPLO 3.20

Para saber ¿cuál es intervalo de clase que contiene al primer cuartil? , se divide la suma de todas las frecuencias entre 4. $\dfrac{\sum_{i=1}^{i=n} fi}{4} = 7.5$ *y al valor resultante se le busca donde está contenido en las frecuencias acumuladas, de igual manera que se procedió al obtener la mediana.*

Int. de clase	*fi*	*fa*	*Xi*	*Int. Reales de clase*
46.30 - 46.52	*4*	*4*	*46.41*	*46.295 - 46.525*
46.52 - 46.75	*4*	*8*	*46.64*	*46.525 - 46.755*
46.76 - 46.98	*16*	*24*	*46.87*	*46.755 - 46.985*
46.99 - 47.21	*4*	*28*	*47.09*	*46.985 - 47.225*
47.22 - 47.44	*2*	*30*	*47.33*	*47.225 - 47.445*

intervalo de clase que contiene al primer cuartil *de donde.*

$$Q_1 = 46\,525 + \frac{7\,5-4}{4} \quad .23$$

$$= 46.525 + (.87)(.23)$$

$$= 46.525 + .2001$$

$$= 46.7251 \approx \mathbf{46.73}$$

Segundo cuartil Q_2.

El valor del segundo cuartil es el mismo que la mediana en cualquier ejercicio que se efectúe, sin embargo, se pondrá su fórmula.

$$Q_2 = L_{Q_2} + \frac{\frac{\sum_{i=1}^{i=n} fi}{2} - (\sum f)_{Q_2}}{f_{Q_2}} \quad c$$

donde·

Q_2 = *Segundo cuartil*

L_2 = *límite real inferior del intervalo que contiene al segundo cuartil.*

$\sum_{i=1}^{i=n} fi$ = *suma de todas las frecuencias*

$(\sum f)_{Q_2}$ = *suma de todas las frecuencias por debajo de la frecuencia del intervalo que contiene al segundo cuartil*

c = *tamaño del intervalo que contiene al segundo cuartil*

nuevamente se requiere encontrar el intervalo que contiene al segundo cuartil, para lo cual, se divide el total de las frecuencias entre dos y el resultado se busca donde está contenido en las frecuencias acumuladas.

$$\frac{\sum_{i=1}^{i=n} fi}{2} = \frac{30}{5} = 15$$

buscando donde está contenido el resultado en las frecuencias acumuladas

Int. de clase	*fi*	*fa*	*Xi*	*Int. Reales de clase*
46.30 - 46.52	*4*	*4*	*46.41*	*46.295 - 46.525*
46.52 - 46.75	*4*	*8*	*46.64*	*46.525 - 46.755*
46.76 - 46.98	*16*	*24*	*46.87*	*46.755 - 46.985*
46.99 - 47.21	*4*	*28*	*47.09*	*46.985 - 47.225*
47.22 - 47.44	*2*	*30*	*47.33*	*47.225 - 47.445*

intervalo de clase que contiene al segundo cuartil

Por lo que.

$$Q_2 = 46\ 755 - \left(\frac{15-8}{16}\right).23$$

$$= \mathbf{46.84}$$

Tercer cuartil Q_3.

$$Q_3 = L_{Q3} + \frac{\frac{3\sum_{i=1}^{i=n} fi}{4} - (\sum f)_{Q3}}{f_{Q3}} c$$

donde

Q_3 = *primer cuartil*

L_3 = *límite real inferior del intervalo que contiene al tercer cuartil.*

$\sum_{i=1}^{i=n} fi$ = *suma de todas las frecuencias.*

$(\sum f)_{Q3}$ = *suma de todas las frecuencias por debajo de la frecuencia del intervalo que contiene al tercer cuartil.*

c = *tamaño del intervalo que contiene al tercer cuartil.*

para encontrar el intervalo de clase que contiene al tercer cuartil, se multiplica la suma de todas las frecuencias por 3 y se divide entre 4 La cantidad resultante se busca en que frecuencia acumulada esta contenida dicha cantidad.

Int. de clase	*fi*	*fa*	*Xi*	*Int. Reales de clase*
46.30 - 46.52	*4*	*4*	*46 41*	*46.295 - 46.525*
46.52 - 46.75	*4*	*8*	*46.64*	*46.525 - 46.755*
46.76 - 46.98	*16*	*24*	*46.87*	*46.755 - 46.985*
46.99 - 47.21	*4*	*28*	*47.09*	*46.985 - 47.225*
47.22 - 47.44	*2*	*30*	*47.33*	*47.225 - 47.445*

intervalo de clase que contiene al tercer cuartil

$$Q_3 = 46.755 + \frac{22.5 - 8}{16} \quad .23$$

$$= \mathbf{46.96}$$

MEDIDAS DE VARIABILIDAD

Al grado en que los datos numéricos tienden a extenderse alrededor de un valor medio se le llama variación o dispersión. Se utilizan distintas medidas de dispersión o variación, las más empleadas son: , ***la desviación media, la desviación estándar y la varianza****. Dichas medidas son parámetros estadísticos que tienen como objetivo principal, evaluar la calidad de lo datos Por ejemplo, si se quisiera hacer un estudio para obtener las características del habitante del Distrito Federal, es decir, características físicas (color de ojos, color de la piel, altura, etc.) y características socioculturales (grado escolar, ocupación, número de hijos , etc.), se tendría que procurar tomar gente que viva en el Distrito Federal para obtener una muestra representativa no sesgada que arroje resultados muy apegados a la realidad. Sin embargo, si extremadamente, se eligiera una muestra de gente de Canadá para dar características de la gente que vive en el Distrito Federal, obviamente no daría resultados confiables, pues son poblaciones de características muy diferentes.*

En el ejemplo que se mencionó, se puede observar a simple vista que se esta tomando un caso extremo, que nunca sucedería, sin embargo, existen en la realidad sesgos que llevan a una mala elección de los datos o muestras, que no son tan evidentes, por lo que las medidas de dispersión se hacen necesarias.

DATOS NO AGRUPADOS

Las medidas de dispersión que se estudiaran en este apartado, son medidas de dispersión con respecto a la media aritmética.

DESVIACIÓN MEDIA

La desviación media o promedio de desviaciones, de una serie de datos X_1, X_2, X_3, ...Xn, viene definido por

$$\text{Desviación Media} = \frac{\sum_{i=1}^{i=n} |Xi - \overline{X}|}{n}$$

donde:

$$Xi = \text{dato iésimo}$$

$$\overline{X} = \text{media aritmética}$$

$$|Xi - \overline{X}| = \text{valor absoluto de las desviaciones}$$

y cuyo desarrollo sería.

$$Desviación\ Media = \frac{\sum_{i=1}^{i=n} |Xi - \overline{X}|}{n}$$

$$= \frac{|X_1 - \overline{X}| + |X_2 - \overline{X}| + |X_3 - \overline{X}| + \quad + |Xn - \overline{X}|}{n}$$

EJEMPLO 3.21

Se tomaron las edades de 10 personas tomadas al azar.

24, 34, 56, 8, 67, 23, 27, 17, 23, 42

y se quiere saber la desviación media, para lo cual es necesario calcular la media aritmética $\overline{X}$.

$$\overline{X} = \frac{24 + 34 + 56 + 8 + 67 + 23 + 27 + 17 + 23 + 42}{10}$$

$$= \frac{321}{10} = \mathbf{32.10}$$

ahora, calculando la desviación media.

$$Desv\ Med = \frac{|24-32\ 10| - |34-32\ 10| - |56-32\ 10| - |8-32\ 10| - |67-32\ 10| - |23-32\ 10|}{10}$$

$$\frac{| - |27-32\ 10| - |17-32\ 10| - |23-32\ 10| - |42-32\ 10|}{}$$

$$= \frac{|-8.1| + |1.9| + |23.9| + |-24.1| + |34.9| + |-9.1| + |-5.1| + |-15.1| + |-9.1| + |9.9|}{10}$$

$$= \frac{8.1 + 1.9 + 23.9 + 24.1 + 34.9 + 9.1 + 5.1 + 15.1 + 9.1 + 9.9}{10}$$

$$= \frac{141.2}{10} = \mathbf{14.12}$$

DESVIACIÓN ESTÁNDAR

La desviación estándar debe ser definida, tomando en cuenta dos aspectos importantes. ***i) Cuando los datos representan una muestra*** *y se define como*

Sean $X_1, X_2, X_3, \ldots Xn$ *una muestra de n datos, entonces.*

$$s = \sqrt{\frac{\sum_{i=1}^{i=n} (Xi - \overline{X})}{n-1}}$$

donde.

s = *desviación estándar para una muestra*

Xi = *dato iésimo*

$\overline{X}$ = *media aritmética*

n = *número de datos*

La desviación estándar de los datos de una muestra se definió con (n-1) en lugar de n, porque así el valor resultante representa un mejor estimador de la *población de la que fue extraída la muestra. Para valores grandes de n (n 30) prácticamente no hay diferencia entre hacer esta transformación o no.*

ii) ***Cuando se toma en cuenta una población****, es representada por σ y su fórmula es.*

$$\sigma = \sqrt{\frac{\sum_{i=1}^{i=n}(Xi - \overline{X})^2}{n}}$$

σ = *desviación estándar para la población*

Xi = *dato iésimo*

$\overline{X}$ = *media aritmética*

n = *número de datos*

EJEMPLO 3.22

Tomando en cuenta los datos del ejemplo 3.19, donde se le preguntó la edad a una muestra de diez personas tomadas al azar.

24, 34, 56, 8, 67, 23, 27, 17, 23, 42

la media aritmética fue.

$$\overline{X} = 32\,10$$

calculando la desviación estándar, como es una muestra, la fórmula será.

$$s = \sqrt{\frac{\sum_{i=1}^{i=n}(Xi - \overline{X})}{n-1}}$$

$$s = \sqrt{\frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \quad + (Xn - \overline{X})^2}{n-1}}$$

sustituyendo

$$s = \sqrt{\frac{(24 - 32.10)^2 + (34 - 32.10)^2 + (56 - 32\,10)^2 + (8 - 32.10)^2 + (67 - 32\,10)^2 +}{10-1}}$$

$$\frac{(23 - 32\,10)^2 + (27 - 32\,10)^2 + (17 - 32\,10)^2 + (23 - 32\,10)^2 + (42 - 32\,10)^2}{10-1}$$

$$s = \sqrt{\frac{65\,61 + 3\,61 + 571.21 + 580\,81 + 1210\,01 + 82\,81 + 26\,01 + 228\,01 + 82\,81 + 98\,01}{9}}$$

$$= \sqrt{\frac{2948\,90}{9}} = \sqrt{327\,65} = 18\,18$$

VARIANZA

Por lo que se refiere a la varianza de un conjunto de datos, sucede lo mismo que con la desviación estándar, en lo que se refiere a los datos poblacionales o muestrales, la pequeña diferencia en sus fórmulas también aquí se hace sentir.

Es importante notar, que la varianza y la desviación estándar son medidas de *dispersión muy semejantes pero distintas, que al obtener una, fácilmente se puede calcular la otra*

La varianza de una muestra, es el valor numérico que se obtiene con la siguiente fórmula.

$$s^2 = \frac{\sum_{i=1}^{i=n}(Xi - \overline{X})^2}{n-1}$$

por lo que se puede observar, la varianza es el cuadrado de la desviación estándar o bien la desviación estándar es la raiz cuadrada de la varianza. Por lo que al ser calculada una, como ya se mencionó, es fácil calcular la otra.

$$\text{Varianza} = (\text{Desviación estándar})^2$$

o bien

$$\text{Desviación estándar} = \sqrt{\text{varianza}}$$

se tenía que:

$$s = \mathbf{18.18}$$

de donde

$$s^2 = (18.18)^2$$

$$= \mathbf{330.51}$$

PARA DATOS AGRUPADOS.

EJEMPLO 3.23

Tomando los datos del ejemplo 3.10. Los siguientes datos, representan una muestra en grados centígrados de una cierta materia.

47.20	46.54	46.42	46.60	46.89	47.40	46.90	46.30
47.40	46.90	47.12	46.88	46.87	46.53	46.93	46.80
47.10	46.69	46.90	47.22	46.96	46.48	47.00	46.54
46.94	46.35	46.87	46.78	46.81	46.90	46.52	46.89

La tabla muestra la organización en cinco intervalos de clase de los datos anteriores.

Intervalos de clase	**conteo**	**frecuencia fi**
46.30 - 46.52	xxxx	4
46.53 - 46.75	xxxx	4
46.76 - 46.98	xxxxxxxxxxxxxxxx	16
46.99 - 47.21	xxxx	4
47.22 - 47.44	xx	2

La varianza y la desviación estándar para datos agrupados, se pueden expresar con las siguientes fórmulas.

VARIANZA ***(para datos muéstrales)***

$$s^2 = \frac{\sum_{i=1}^{i=n} fi(Xi - \overline{X})^2}{\sum_{i=1}^{i=n} fi - 1}$$

londe.

s^2 = varianza para una muestra

X_i = dato iésimo

$\overline{X}$ = media aritmética

$\sum_{i=1}^{i=n} fi$ = suma de todas las frecuencias

para calcular la varianza, es necesario calcular antes, la marca de clase X_i, la media aritmética y las desviaciones, que serán calculadas en el siguiente cuadro.

Se sabe que:

$$\overline{X} = \mathbf{46.84}$$

de donde

Int. de clase	fi	Xi	$Xi - \overline{X}$	$(Xi - \overline{X})^2$	$fi(Xi - \overline{X})^2$
46.30 - 46.52	4	46.41	-0.43	.1849	.7396
46.52 - 46.75	4	46.64	-0.20	.0400	.1600
46.76 - 46.98	16	46.87	0.03	.0009	.0144
46.99 - 47.21	4	47.09	0.25	.0625	.2500
47.22 - 47.44	2	47.33	0.49	.2401	.4802

$$\boxed{\sum_{i=1}^{i=n} fi = 30} \qquad \boxed{\sum_{i=1}^{i=n} fi(Xi - \overline{X})^2 = 1.6442}$$

por lo que

$$s^2 = \frac{16442}{30-1} = \frac{16442}{29} = \mathbf{.0566}$$

DESVIACIÓN ESTÁNDAR ***(para datos muéstrales)***

$$s = \sqrt{\frac{\sum_{i=1}^{i=n} fi(Xi - \overline{X})^2}{\sum_{i=1}^{i=n} fi - 1}}$$

donde:

s = *desviación estándar para una muestra*

Xi = *dato iésimo*

$\overline{X}$ = *media aritmética*

$\sum_{i=1}^{i=n} fi$ = *suma de todas las frecuencias*

como la desviación estándar es igual a la raíz cuadrada de la varianza, entonces.

$$s^2 = .0566$$

$$\therefore$$

$$s = \sqrt{0566} = \mathbf{.2379}$$

Para cuando se toman datos poblacionales, las fórmulas vienen expresadas como:

VARIANZA **(para datos poblacionales)**

$$\sigma^2 = \frac{\sum_{i=1}^{i=n} fi(Xi - \overline{X})^2}{\sum_{i=1}^{i=n} fi}$$

DESVIACIÓN ESTÁNDAR **(datos poblacionales)**

$$\sigma = \sqrt{Var}$$

COEFICIENTE DE VARIACIÓN Y REGLA EMPÍRICA.

Las medidas de dispersión como antes ya se mencionó, miden la variabilidad que existe de un conjunto de datos con respecto a la media aritmética, y pueden en la medida de sus posibilidades evaluar la calidad de los datos, sin embargo, el teorema de Tchebycheff o Chebyshev, La regla empírica y el coeficiente de variación ayudan a comprender mejor el significado de variabilidad

TEOREMA DE TCHEBYCHEFF. **La proporción de cualquier distribución situada dentro de k desviaciones estándar de la media es, por lo menos, 1 - $(\frac{1}{k^2})$, donde k es cualquier número positivo mayor que 1.**

*En otras palabras, si $\overline{X}$ y s (o bien μ y σ) son la media y la desviación estándar de un conjunto de datos y k $\geq$ 1 entonces, el intervalo $(\overline{X} - ks, \overline{X} + ks)$ o $(\mu - k\sigma, \mu + k\sigma)$ contiene como mínimo la (1 - $\frac{1}{k^2}$) * 100% del total de los datos. Por* *ejemplo, si k = 2, el intervalo* $(\overline{X}-2s,\overline{X}+2s)$ *tiene como mínimo al*

$$(1-\frac{1}{k^2})*100\% = 75\%.$$

EJEMPLO 3.24

Con motivo de los festejos del dia del niño, el departamento de relaciones públicas de una fábrica desea conocer el número de hijos que tienen los 200 obreros que ahí laboran. Supóngase que se entrevistaron a todos los obreros, según el orden que tienen en la nómina, se calcularon la media aritmética, la varianza y la desviación estándar, obteniéndose los siguientes resultados.

$$\mu = 3.415$$

$$\sigma^2 = 4.7427$$

$$\sigma = 2.1777$$

El teorema establece que, siempre habrá al menos un 75% de los datos dentro de dos desviaciones estándar de la media (k = 2), es decir.

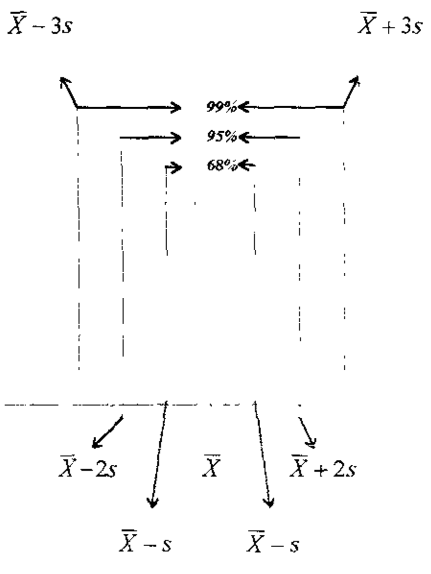
$$1-\frac{1}{k^2} = 1-\frac{1}{4} = \frac{3}{4} = .75 = 75\%$$

así, para el ejemplo anterior, haciendo k=2 el intervalo $(\mu-2\sigma,\mu-2\sigma)$, *sustituyendo, (-.9404, 7.7704) contiene por lo menos el 75% de todos los datos. Si se considera el intervalo que se forma cuando k = 3, es decir, 3 desviaciones estándar en ambos lados de la media, según el teorema se encontrarán por lo menos* $\frac{8}{9}=.8888=88.88\%$ *de los datos. Es obvio que se le pueden asignar valores positivos mayores de 3 calcular el porcentaje de datos que quedan dentro de los intervalos formados, que será más cercana al 100% entre más grande sea el valor k.*

REGLA EMPÍRICA

La regla empírica, establece porcentajes mayores que el teorema de Tchebycheff para cuando k = 2 y K = 3, si y solo si, los datos se aproximan a una forma acampanada. Cuando se esta haciendo estadística descriptiva como es el caso, si el histograma toma esta forma, la regla empírica es una buena herramienta, y se cumple lo siguiente:

En el intervalo que se forma la restarle y sumarle a la media aritmética una desviación estándar, se encuentran aproximadamente el 68% de los datos; Para cuando el intervalo formado es con dos desviaciones es decir (X - 2S, X - 2S) contiene aproximadamente el 95%; Por último, para cuando el intervalo se forma tomado en cuenta tres desviaciones, se aproxima al 99% de los datos, como se puede ver en la siguiente gráfica.



Definición.

Si una variable esta distribuida normalmente, entonces hay un 68% de los datos, aproximadamente, dentro de una desviación estándar de la media Dentro de dos desviaciones estándares hay aproximadamente el 95% y dentro de tres cerca del 99 7%. Esta regla es aplicable específicamente a una distribución normal, la cual se estudiará con detenimiento en el apartado de distribuciones de probabilidad.

EJEMPLO 3.25

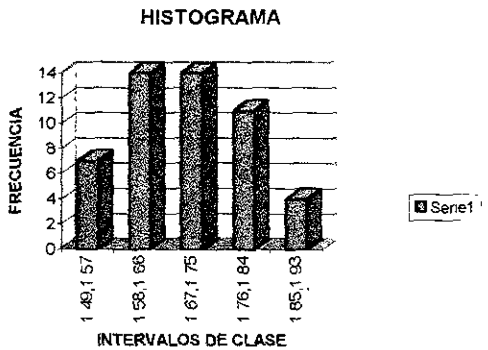
Supónganse los datos del ejemplo 3.9

1.62	1.54	1.91	1 70	1.70
1.80	1.68	1.81	1.85	1.78
1.92	1.69	1.72	1.59	1.50
1.56	1.68	1.70	1.64	1.87
1.58	1.79	1.75	1.73	1.80
1.49	1.49	1.80	1 73	1.60
1.82	1.87	1.53	1.80	1.65
1.63	1.58	1.59	1.62	1.71
1.66	1.69	1.60	1.67	1.59
1.73	1.79	1.63	1.50	1.80

los cuales fueron ordenados en cinco intervalos de clase, que se muestran en la siguiente tabla.

Intervalo de clase	*conteo*	*frecuencia fi*
1.49 - 1.57	xxxxxxx	7
1.57 - 1.66	xxxxxxxxxxxxxx	14
1.67 - 1.75	xxxxxxxxxxxxxx	14
1.76 - 1.84	xxxxxxxxxxx	11
1.85 - 1.93	xxxx	4

y cuyo histograma fue:



se puede observar a simple vista, que el histograma tiende a una forma acampanada, por lo que se puede aplicar la regla empírica Calculando la media aritmética y la desviación estándar, se tiene.

Intervalo de clase	*fi*	*Xi*	*fiXi*	$Xi-\overline{X}$	$(Xi-\overline{X})^2$	$(Xi-\overline{X})^2 fi$
1.49 - 1.57	*7*	*1.53*	*10.71*	*-.16*	*.0256*	*.1792*
1.58 - 1.66	*14*	*1.62*	*22.68*	*-.07*	*.0049*	*.0686*
1.67 - 1.75	*14*	*1.71*	*23.94*	*.02*	*.0004*	*.0056*
1.76 - 1.84	*11*	*1.80*	*19.80*	*.11*	*.0121*	*.1331*
1.85 - 1.93	*4*	*1.89*	*7.56*	*.20*	*.0400*	*.1600*
	50		*84.69*			*5465*

de donde:

$$\overline{X} = \frac{\sum_{i=1}^{i=n} fiXi}{\sum_{i=1}^{i=n} fi} = \frac{84\,69}{50} = \mathbf{1.69}$$

$$s = \sqrt{\frac{\sum_{i=1}^{i=n} fi(Xi - \overline{X})^2}{\sum_{i=1}^{i=N} fi - 1}} =$$

$$= \sqrt{\frac{.5465}{50}} = \sqrt{01093} = 1045$$

obteniendo el intervalo para una desviación estándar.

(1.69 - 1045 , 1.69 – .1045)

(1.70 , 1.79) intervalo

los datos que entran en este intervalo son·

1.62, 1.70, 1.70, 1.68, 1.78, 1.69, 1.59, 1.68, 1.70, 1.64,

1.79, 1.75. 1.73. 1.73, 1.60, 1.65, 1.63, 1.59, 1.62, 1 71,

1.66, 1.69, 1.60, 1.67, 1.59, 1 73, 1.79, 1.63

dando un total de 28, lo que significa que el 56% de los datos se encuentran en este intervalo, si el ejercicio se hiciera con más datos dando por lo tanto un histograma más fino y por lo tanto más tendiente a una forma acampanada, la regla empírica se apega mucho al resultado previsto del 68%.

COEFICIENTE DE VARIACIÓN

Las medidas de dispersión expresadas en valores absolutos, son convenientes para describir la dispersión de un sólo conjunto de valores. Si dos conjuntos de valores están siendo comparados, los valores absolutos son convenientes solamente cuando los promedios de los dos conjuntos son aproximadamente del mismo tamaño y las unidades de medida de los conjuntos son iguales. Es obvio que la comparación de dos diferentes unidades, tales como el número de kilómetros comparados con el número de pesos, no tiene sentido.

Cuando los promedios son claramente diferentes, aunque las unidades pueden ser las mismas, la tarea de comparar los grados de dispersión basada en los valores absolutos de los diferentes conjuntos es aún más difícil. Por ejemplo los pesos de estudiantes del CCH son generalmente mayores que los pesos de estudiantes de escuelas primarias. El promedio y la desviación estándar de los pesos de un grupo de estudiantes del Colegio de Ciencias y Humanidades, por lo tanto se espera que sea mayores que los de cualquier escuela primaria. Por ejemplo, supónganse los siguientes datos.

EJEMPLO 3.26

Comparación de las desviaciones estándares en valores absolutos de 3 alumnos del CCH y 3 alumnos de una escuela primaria

peso en Kgs X	x	x^2	peso en Kgs X	x	x^2
64.5	- 4	.16	17.3	- 5 16	26 65
58 3	- 6 6	43 56	28 4	5 94	35 28
72 0	7 1	50.41	21 4	- 76	57
194 8		94.13	67 4		62 5

Donde $x = (X_i - \overline{X})$ *y el calculo de la media aritmética y desviación estándar* *por grupo son:*

Para alumnos de Colegio de Ciencias y Humanidades.

$$\overline{X} = \frac{194\ 8}{3} = 64\ 9$$

y la desviación estándar.

$$s = \sqrt{\frac{94\ 13}{3}} = \sqrt{31.37} = 5\ 60kgs$$

para alumnos de primaria.

$$\overline{X} = \frac{67\ 4}{3} = 22\ 46$$

y la desviación estándar.

$$s = \sqrt{\frac{62\ 5}{3}} = \sqrt{20\ 3} = 4\ 56$$

Comparando los promedios o medias aritméticas de los dos conjuntos de datos expuestos en la tabla anterior, se puede observar claramente que los valores son diferentes, 64 90 para alumnos del CCH y 22.46 para alumnos de escuela primaria No se puede concluir que la más alta desviación estándar 5.60 , proporcione el más alto grado de dispersión Una medida de dispersión expresada en valores relativos es por tanto, requerida para este tipo de comparación. En general, una dispersión relativa es *el cociente de una medida dada de dispersión dividida por el promedio o media aritmética con respecto al cual las desviaciones fueron medidas.*

La medida de dispersión más comúnmente usada expresada en valores relativos es el ***coeficiente de variación*** *, representado pro V. Es el cociente de la desviación estándar dividida por la media aritmética.*

$$V = \frac{s}{\overline{X}}$$

Calculando esta medida para los pesos de los estudiantes del CCH y los alumnos de la escuela primaria.

$$V = \frac{5.60}{64.9} = 086 = 8.62\% \text{ para los primeros}$$

y

$$V = \frac{4.56}{22.46} = 2030 = 20.30\% \text{ para los segundos}$$

con lo que al comparar estos coeficientes se ve que el coeficiente para alumnos del Colegio es menor que el de los alumnos de escuela primaria o en otras palabras la dispersión relativa de los estudiantes del CCH en este caso es menor. Lo que da como conclusión que no necesariamente a peso mayor de mayor grado de dispersión.

4.- PROBABILIDAD

Se ha estudiado hasta aquí, lo que es estadística descriptiva y la forma en como ésta puede ser utilizada como herramienta en situaciones reales, se ha visto ademas, como los datos son recopilados, organizados y analizados. Sin embargo, resulta indispensable tratar de obtener más y mejores resultados, que permitan no sólo describir las características de una muestra o población, sino hacer inferencia sobre ellas. Es por tal motivo, que el conocimiento de la Probabilidad es de suma importancia en todo estudio estadístico.

Cabe destacar que, para lograr mejores resultados en el estudio de la Probabilidad, es necesario tener buenas bases en lo que se refiere a la teoría de conjuntos y el concepto de factorial.

4.1 PROBABILIDAD CLÁSICA Y FRECUENCIAL

La familiaridad con la que mucha gente maneja el concepto de probabilidad es alta, muchas veces se ha oído decir ¿cuál es la probabilidad de que me saque la lotería o el me late? o ¿qué posibilidad hay de que me pase una accidente automovilístico? o ¿qué posibilidad hay de que hoy llueva? Para ver si me llevo paraguas o no, o ¿qué posibilidad existe que me reprueben en matemáticas?, etc.. Existen muchas más, que sin saber que se están haciendo preguntas sobre probabilidad, son cotidianas. Es cierto que, otros sectores de la población, que por el trabajo que realizan, tienen la necesidad de conocer la probabilidad, como es el caso de las compañías de seguros, donde se maneja el riesgo en forma probabilística para asignar una cierta cantidad llamada la prima, otro ejemplo, es la gente que se dedica a los juegos de azar, como es la lotería nacional o el me late. Así como, empresas públicas y privadas que necesitan la elaboración de estudios estadísticos inferenciales, un ejemplo de esto es le INEGI, Instituto Nacional de Estadística, Geografía e Informática.

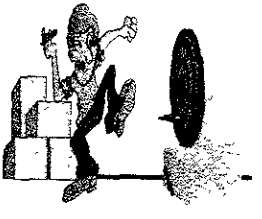
Con anterioridad se definieron los conceptos de fenómeno determinista y fenómeno aleatorio y se dijo también, que en estos apuntes se cambiaría el concepto de fenómeno por el de experimento.

Se sabe que la probabilidad es el estudio de fenómenos aleatorios o libres de determinación, sin embargo, para definir lo que significa probabilidad clásica y probabilidad frecuencial, es necesario saber que se entiende por Espacio Muestral y por Evento.

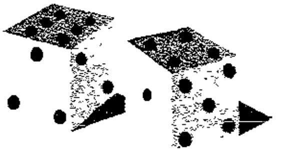
Experimento.- *Es aquella acción que se considera con propósito de análisis y que tiene como fin último determinar la probabilidad de uno o de varios resultados.*

EJEMPLO 4.1

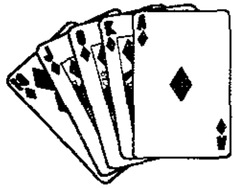
i) Tirar dardos a un blanco determinado



ii) Lanzar un par de dados



iii)Obtener una carta de una baraja americana



ESPACIO MUESTRAL S .- es el conjunto de todos los posibles resultados de interés de un experimento dado.

EJEMPLO 4.2

i) Experimento.- Se lanza una moneda.

El espacio muestral sería el total de formas en como puede caer la moneda. En este caso solo existen dos formas de interés, que caiga sol o que caiga águila Sin embargo, se han dado casos en los cuales, puede caer la moneda de canto, no obstante como esto no representa interés alguno se procederá a lanzar la moneda nuevamente Así el espacio muestral, sería:

$$S = \{s, a\}$$

ii) Experimento. Se lanza un dado.

Obviamente aquí sucede algo parecido a lo anterior, que el dado pueda caer en una en una punta, sin que caiga ninguna cara de interés, por tal motivo, igual que en el caso de la moneda se procederá a lanzar nuevamente el dado tantas veces como sea *necesario para obtener un resultado de interés. El espacio muestral es:*

$$S = \{1,2,3,4,5,6\}$$

***Evento.-** es un subconjunto de resultados posibles considerados previamente en el espacio muestral y se denota comúnmente con las letras mayúsculas A, B, C.*

EJEMPLO 4.3

i) Sea A el evento de que al nacer un bebe, éste sea niña.
ii) Sea B el evento de que una persona de 20 años de edad, sobreviva 15 años más.
iii) Sea C el evento de que al extraer una carta de una baraja española, salga un rey.

Evento Seguro

Cuando los elementos de un evento cualquiera son los mismos del espacio muestral, se llama evento seguro.

Evento Imposible

Cuando un evento no tiene elementos de interés para un fenómeno, se llama evento imposible.

Cabe destacar, cuando se estudia probabilidad suelen utilizarse los juegos de azar para su mejor compresión, no obstante se tratará de hacer algunos ejercicios más apegados a la realidad.

Definición (Probabilidad Frecuencial y Regularidad Estadística)

*Las frecuencias relativas de un evento tienden a estabilizarse cuando el número de observaciones del fenómeno se hace cada vez mayor. **La regularidad estadística,** en* *el experimento del lanzamiento de monedas, indica que las frecuencias relativas del evento sol $\{s\}$ se tienden a estabilizar aproximadamente en .5 con lo que se puede proporcionar la siguiente definición.*

La probabilidad de un evento A, denotada por P(A) es el valor en el que se estabilizan las frecuencias relativas del evento A cuando el número de observaciones del experimento se hace cada vez mayor.

Esta definición proporciona una forma útil para asignar probabilidad a los eventos, aunque en ocasiones es difícil de saber el valor para el cual se estabilizan las frecuencias relativas del evento, debido a que el experimento no es observado las veces necesarias para obtener una regularidad estadística.

Definición (Probabilidad Clásica)

Sea S un espacio muestral cualquiera y A un evento de ese espacio. Se define la probabilidad P del evento A, como:

$$P(A)=\frac{\text{número de casos favorables}}{\text{número de casos totales}}$$

en otras palabras

$$P(A)=\frac{\text{número de elementos del evento } A}{\text{número de elementos del espacio muestral } S}$$

$$P(A)=\frac{n(A)}{n(S)}$$

donde $n(A)$ y $n(S)$ se conocen como la cardinalidad de A y de S respectivamente y consiste en el número de elementos de ambos.

EJEMPLO 4.4

Experimento. *Se lanza una moneda.*

Sea A el evento de que al lanzar una moneda, caiga águila. Calcular al probabilidad de A.

Se sabe que por definición.

$$P(A) = \frac{n(A)}{n(S)}$$

por tal motivo se debe calcular el número de elementos del espacio muestral $n(S)$ y posteriormente el número de elementos del evento $n(A)$.

El espacio muestral es:

$$S = \{ s,a \} \qquad n(S) = 2$$

por otro lado

$$A = \{ s \} \qquad n(A) = 1$$

por lo tanto, la probabilidad de A es.

$$P(A) = \frac{1}{2}$$

EJEMPLO 4.5

Experimento. *Se lanza un dado.*

En este experimento, se pueden definir más eventos, por ejemplo

- *Sea A el evento de que al lanzar un dado, caiga un número par.*
- *Sea B el evento de que al lanzar un dado caiga un número mayor de cuatro*
- *Sea C el evento de que al lanzar un dado caiga dos*

La pregunta obligada sería ¿Cuál es la probabilidad de A, B y C?. Obteniendo el conjunto de todos los posibles resultados es decir el espacio muestral S y cada uno de los eventos.

$$S = \{1,2,3,4,5,6\}$$

el número de elementos que pertenecen al espacio muestral es :

$$n(S) = 6$$

así mismo, el evento A quedaría representado como.

$$A = \{2,4,6\}$$

el número de elemento que pertenecen al evento A es

$$n(A) = 3$$

$$P(A) = \frac{n(A)}{n(S)} = \frac{3}{6}$$

para el evento B

$$B = \{5,6\}$$

y cuyo número de elementos

$$n(B) = 2$$

$$P(B) = \frac{n(B)}{n(S)} = \frac{2}{6} = \frac{1}{3}$$

por lo que corresponde al evento C.

$$C = \{2\}$$

de donde

$$n(C) = 1$$

obteniendo probabilidades por definición.

$$P(C) = \frac{n(C)}{n(S)} = \frac{1}{6}$$

EJEMPLO 4.6

Experimento. *Se lanzan dos monedas.*

En este caso, el espacio muestral consta de cuatro resultados posibles

$$S = \{ss, sa, as, aa\} \quad ; \quad n(S) = 4$$

existen tres conceptos que se manejan en probabilidad y que se usan de manera cotidiana, el primero sería el concepto de ***exactamente****, que significa como su nombre lo dice la exactitud del evento; el segundo,* ***por lo menos,*** *cuyo significado es la cantidad que se mencione o más; el tercero y ultimo* ***a lo más,*** *que se refiere al valor que se mencione en el evento o menos. El siguiente ejemplo trata de esclarecer estos tres conceptos.*

- *Sea A el evento de que al lanzar dos monedas caiga un sol* ***exactamente.***
- *Sea B el evento de que al lanzar dos monedas caiga* ***por lo menos*** *un sol.*
- *Sea C el evento de que al lanzar dos monedas caiga* ***a lo más*** *uno*

Obteniendo las probabilidades de cada uno de los eventos

$$A = \{sa, as\} \quad ; \quad n(A) = 2$$

como se puede observar, los elementos del evento A tienen la característica de un sólo sol, aún cuando se completen con águilas. De donde.

$$P(A) = \frac{n(A)}{n(S)} = \frac{2}{4}$$

Por lo que se refiere a B. Los elementos de este evento, tendrán la característica de encontrar en ellos un sol o más de un sol.

$$B = \{sa, as, ss\} \quad ; \quad n(B) = 3$$

$$P(B) = \frac{n(B)}{N(S)} = \frac{3}{4}$$

por ultimo, para el evento C (a lo más), los elementos de este evento, tendrán la caracteristica de tener un sol o ningún sol

$$C = \{sa, as, aa\} \quad ; \quad n(C) = 3$$

$$\boxed{P(C) = \frac{n(C)}{n(S)} = \frac{3}{4}}$$

EJEMPLO 4.7

Experimento. *Se lanzan tres monedas.*

Es obvio, que cuando el número de monedas que son lanzadas es mayor, obtener el total de resultados posibles y visualizarlos en forma de conjunto es tambien más complejo. Es por tal motivo, que se hace necesaria la introducción de una técnica de conteo llamada el ***Principio Fundamental del Conteo,*** *y que se estudiará con más profundidad en el apartado correspondiente*

PRINCIPIO FUNDAMENTAL DEL CONTEO

Si en un experimento, un primer ensayo puede ocurrir de n_1 maneras diferentes y si continuando el experimento, un segundo ensayo puede ocurrir de n_2 maneras diferentes, y si continuando el experimento un tercer ensayo puede ocurrir de n_3 maneras diferentes, y así sucesivamente, entonces el número de formas como puede ocurrir el experimento es igual a $n_1 + n_2 + n_3 \ldots$

Es importante aclarar el significado de ensayo dentro del principio fundamental *del conteo, y éste corresponde a cada repetición que se realice en un experimento. Por ejemplo, cuando el experimento consiste en el lanzamiento de monedas, cada moneda corresponde a un ensayo, así cuando es el caso sea lanzamiento de dados, cada dado será el equivalente a un ensayo.*

El principio fundamental del conteo como una técnica para contar, es muy flexible, por lo que se usa con mucha frecuencia, pues sus alcances son muy extensos.

Continuando con el ejemplo 4.7 (se lanzan tres monedas), aplicando el principio fundamental del conteo, se tienen en este caso tres ensayos que corresponderían como antes se mencionó a las tres monedas. Obteniendo el número de resultados posibles que no es otra cosa que el número de elemento del espacio muestral. Se tiene:

$$n(S) = \underset{n_1}{\underline{2}}\ \underset{n_2}{\underline{2}}\ \underset{n_3}{\underline{2}} = 8$$

por lo que.

$$n(S) = 8$$

Así mismo, existe un manera sencilla de acomodar los elementos del espacio muestral, siempre y cuando solo tenga dos posibles resultados (águila - sol, verdadero - falso, hombre o mujer, etcétera). Esta forma no tiene un nombre específico, pero en estos apuntes será definida como **El algoritmo de la mitad de las mitades.** *Esto es, porque al aplicar el algoritmo, al total de resultados, que en este ejemplo son ocho, se le irá dividiendo entre dos, es decir, si se divide ocho entre dos, es igual a cuatro, el método consiste en poner cuatro soles y cuatro águilas en la primera columna, posteriormente, cuatro se debe dividir entre dos, resultando dos. Por lo que hay que poner dos soles y* *dos águilas, hasta completar la segunda columna, por último, dos se divide entre dos, resultando uno, por lo que la tercer columna debe ser completada con un sol y una águila. Se debe tener cuidado de comenzar siempre por soles o siempre por águilas para que el algoritmo funcione.*

Algoritmo de la mitad de las mitades

$8/2 = 4$	$4/2 = 2$	$2/2 = 1$
1^a columna	2^a columna	3^a columna
s	s	s
s	s	a
s	a	s
s	a	a
a	s	s
a	s	a
a	a	s
a	a	a

por lo que los elementos del espacio muestral son:

$$S = \{sss,\ ssa,\ sas,\ saa,\ ass,\ asa,\ aas,\ aaa\}$$

- *Sea A el evento de que al lanzar tres monedas caigan dos soles exactamente.*

¿cuál es la probabilidad de que suceda A, $P(A)$ *?.*

$$A = \{saa,\ asa,\ aas\}$$

el número de elementos de A es igual a tres. $n(A) = 3$

por lo tanto, por definición.

$$\boxed{P(A) = \frac{n(A)}{n(S)} = \frac{3}{8}}$$

- *Sea B el evento de que al lanzar tres monedas caiga a lo más un sol.*

$$B = \{saa, asa, aas, aaa\}$$

el número de elementos de B es igual a cuatro. $n(B) = 4$

de donde

$$\boxed{P(B) = \frac{n(B)}{n(S)} = \frac{4}{8} = \frac{1}{2}}$$

EJEMPLO 4.8

Experimento. *Se lanzan dos dados.*

Según la definición del principio fundamental del conteo, en este experimento habría dos ensayos, de donde

$$n(S) = \underset{n_1}{6} \quad \underset{n_2}{6} = 36$$

$$S = \begin{matrix} 11, & 21, & 31, & 41, & 51, & 61 \\ 12, & 22, & 32, & 42, & 52, & 62 \\ 13, & 23, & 33, & 43, & 53, & 63 \\ 14, & 24, & 34, & 44, & 54, & 64 \\ 15, & 25, & 35, & 45, & 55, & 65 \\ 16, & 26, & 36, & 46, & 56, & 66 \end{matrix}$$

- *Sea A el evento de que al lanzar un par de dados, la suma de los puntos sea 7 exactamente.*

$$A = \{16, 25, 34, 43, 52, 61\} \quad ; \quad n(A) = 6$$

por lo tanto.

$$P(A) = \frac{6}{36} = \frac{1}{6}$$

- *Sea B el evento de que al lanzar un par de dados, la suma de sus puntos sea mayor que nueve.*

$$B = \{46, 55, 64, 56, 65, 66\} \quad , \quad n(B) = 6$$

por lo que

$$P(B) = \frac{n(B)}{n(S)} = \frac{6}{36} = \frac{1}{6}$$

$$n(A) = \underline{9}\ \ \underline{9}\ \ \underline{4} = 324$$
$$\qquad\quad n_1\ \ n_2\ \ n_3$$

por lo tanto:

$$\boxed{P(A) = \frac{n(A)}{n(S)} = \frac{324}{729}}$$

4.2 LEYES DE LA PROBABILIDAD

4.2.1 Evento Seguro y Evento Imposible.

Como antes se mencionó, el espacio muestral S es el conjunto de todos los posibles resultado de interés de un experimento dado y evento es un subconjunto de resultados posibles considerados previamente en el espacio muestral. Si un evento A consta de un sólo elemento $A = \{a\}$, $a \in S$ se llama evento elemental. Así mismo, $\varnothing$ conjunto vacío o evento sin elementos, llamado evento imposible y S evento cierto o seguro.

Cuando se estudian las leyes de probabilidad o probabilidad axiomática, es necesario tener en cuenta las operaciones que se pueden hacer entre eventos y lo que se entiende por eventos mutuamente exclusivos o disjuntos Así como, lo que se entiende por eventos independientes.

Las operaciones que se pueden hacer con eventos son.

1. *$A \cup B$ llamada la unión de los eventos originales, es el evento que sucede, si y solo si A sucede o B sucede o ambos suceden.*
2. *$A \cap B$ llamada la intersección de los eventos originales , es el evento que sucede, si y solo si A y B suceden simultáneamente*

3. A^c llamado el complemento del evento A, es el evento que sucede, si y solo si no sucede A

Las relaciones que se dan entre los eventos al ser aplicadas estas operaciones, pueden comprenderse mejor y facilitarse haciendo uso de los axiomas y teoremas de probabilidad (Leyes de la Probabilidad), que se verán a continuación.

Antes de comenzar de debe entender perfectamente bien que se entiende por axioma y que por teorema.

Axioma.- *Es una verdad tan evidente que no requiere de demostración*

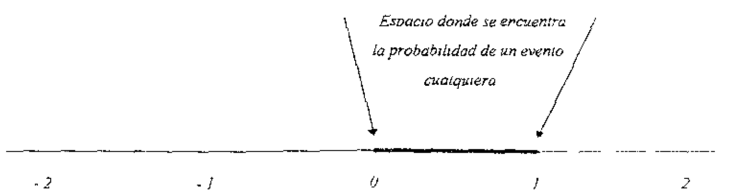
Teorema.- *Es una verdad que requiere ser demostrada*

AXIOMA 1

Sea S un espacio muestral cualquiera y sea A un evento, tal que $A \subset S$, entonces se cumple que:

$$0 \leq P(A) \leq 1$$

es decir, la probabilidad de cualquier evento no puede ser más grande que uno ni ser menor que cero. Cuando es igual a 1 se llama certeza y cuando es cero se llama evento imposible. Supóngase un recta real.



AXIOMA 2

$$P(S) = 1$$

Cuando la probabilidad de un evento vale 1, dicho evento es llamado evento seguro, esto quiere decir, que si el evento ocurre, su ocurrencia será totalmente segura. Obsérvese que el espacio muestral S siempre será un evento seguro.

EJEMPLO 4.10

Experimento. *Se lanza un dado.*

De acuerdo a la definición de probabilidad clásica de un evento A antes considerada, se tiene que :

$$P(A) = \frac{n(A)}{n(S)}$$

entonces, si $A = S$, esto es, si el evento A coincide o es igual al espacio muestral, entonces.

$$P(S) = \frac{n(S)}{n(S)} = 1 \quad \textbf{\textit{la certeza}}$$

De una manera análoga pero totalmente opuesta, se puede definir $\varnothing$ como el evento imposible, como aquel cuya probabilidad de ocurrencia es igual a cero. Por ejemplo.

- *Una persona que quiere ganar la lotería nacional, pero no compra boleto.*
- *El evento de que aparezca siete al lanzar un dado.*
- *El evento de que una persona viva 200 años*

Se dice que en estos casos, los eventos son vacíos, por lo tanto, se puede considerar el siguiente resultado:

TEOREMA 1

Si $\varnothing$ es el conjunto vacío, entonces la probabilidad de que $\varnothing$ es igual a 0

$$P(\varnothing) = 0$$

4.2.2 UNIÓN, INTERSECCIÓN Y COMPLEMENTO DE EVENTOS.

UNIÓN

La unión de dos eventos, se presenta de dos formas diferentes. Cuando los eventos son ***mutuamente exclusivos*** *(que no tienen elementos en común) y cuando entre los eventos hay elementos comunes.*

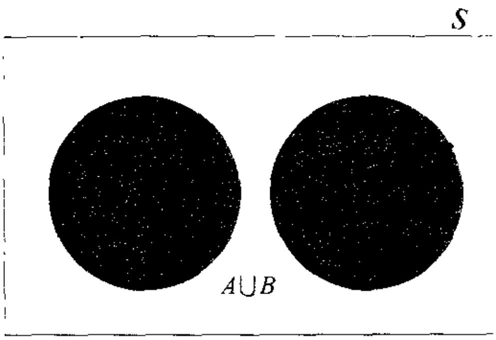
Definición.

Se dice que dos eventos A y B son mutuamente exclusivos, cuando no pueden ocurrir simultáneamente, es decir, $A \cap B = \varnothing$.

Tomado en primera ins . ncia, la unión para dos eventos mutuamente exclusivos, surge el siguiente axioma.

AXIOMA 3

Sea S un espacio muestral cualquiera y sean A y B dos eventos tales que $A \subset S$, $B \subset S$ y $A \cap B = \varnothing$, es decir, dos eventos mutuamente exclusivos.



entonces.

$$P(A \cup B) = P(A) - P(B)$$

en otra palabras, la probabilidad de dos eventos mutuamente exclusivos es igual a la suma de sus probabilidades.

EJEMPLO 4.11

Experimento. *Se lanzan dos monedas.*

El espacio muestral correspondiente está dado por.

$$S = \{ss, sa, as, aa\}$$

- *Sea A el evento de que al lanzar un par de monedas, caigan dos soles exactamente*
- *Sea B el evento de que al lanzar un par de monedas, caiga un sol exactamente.*

Los elementos de cada evento, son

$$A = \{ss\} \quad y \quad B = \{sa, as\}$$

como fácilmente se puede observar, $A \cap B = \varnothing$ *pues* $\{ss\} \cap \{sa, as\} = \varnothing$ *esto es no hay elementos comunes entre ambos conjuntos por lo que, los eventos son mutuamente exclusivos o disyuntos, de tal manera que.*

$$P(A \cup B) = P(A) + P(B) = \frac{1}{4} + \frac{2}{4} = \frac{3}{4}$$

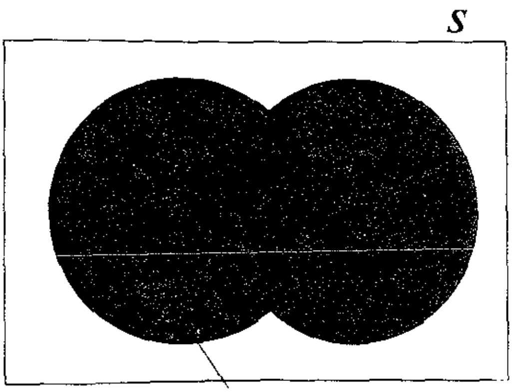
AXIOMA 4

Sean A_1 A_2 A_3, A_4, . A_n *eventos mutuamente exclusivos.*

$$P(A_1 \cup A_2 \cup A_3 \cup A_4 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + P(A_3) + P(A_4), \ldots + P(A_n)$$

la probabilidad de varios eventos mutuamente exclusivos, es igual a la suma de sus probabilidades.

Tomando en cuenta el segundo caso, ***los eventos A y B no son disyuntos,*** *es decir, tienen elementos comunes.* $A \cap B \neq \varnothing$. *Se da el siguiente teorema.*



$A \cup B \neq \varnothing$

TEOREMA 2

Sean A y B dos eventos no disyuntos o lo que es lo mismo $A \cap B \neq \varnothing$, *entonces*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

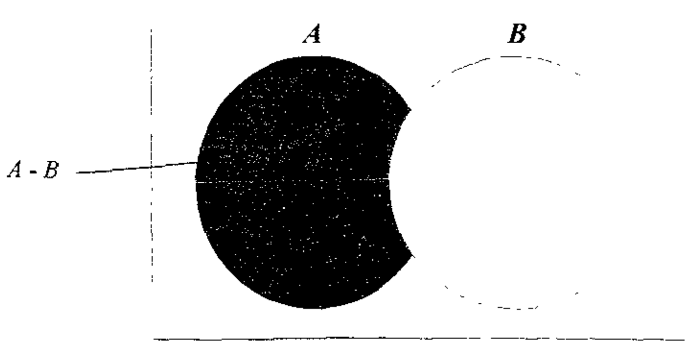
Demostración

Se sabe que:

$$A \cup B = \{x \,/\, x \in A \text{ ó } x \in B\} \quad y \quad A \cap B \neq \phi \quad \text{en este caso.}$$

Por lo que los elementos de la intersección pertenecen a A y también pertenecen a B y

$$A - B = \{x \,/\, x \in A \,,\ x \notin (A \cap B)\}$$



por lo que $A \cup B$ *se puede descomponer en los eventos A - B y B que son* *mutuamente exclusivos, esto es* $A \cup B = (A - B) \cup B$ *, entonces por el axioma 3 y la definición de A - B.*

$$P(A \cup B) = P(A - B \quad P(B)$$

$$= P(A) - P(A \cap B) - P(B)$$

$$P(A \cup B) = P(A) - P(B) - P(A \cap B)$$

EJEMPLO 4.12

Se lanza un dado y una moneda.

Obteniendo los elementos del espacio muestral correspondiente, se tiene que éste consiste en:

$$S = \{1s,\ 2s,\ 3s,\ 4s,\ 5s,\ 6s,\ 1a,\ 2a,\ 3a,\ 4\ a,\ 5a,\ 6a\}$$

- *Sea A el evento de que al lanzar un dado y una moneda aparezca el número dos o tres con sol*
- *Sea B el evento de que al lanzar un dado y una moneda caigan números pares con sol*

$$A = \{2s,\ 3s\} \qquad B = \{2s,\ 4s,\ 6s\} \qquad A \cap B = \{2s\}$$

como se puede ver, en los eventos anteriores hay elementos comunes, por lo que no son mutuamente excluyentes , en consecuencia la probabilidad en cuestión está dada por.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$= \frac{n(A)}{n(S)} + \frac{n(B)}{n(S)} - \frac{n(A \cap B)}{n(S)}$$

$$= \frac{2}{12} + \frac{3}{12} - \frac{1}{12} = \frac{4}{12} = \frac{1}{3}$$

obsérvese que

$$A \cup B = \{2s,\ 3s,\ 4s,\ 6s\}$$

y que

$$P(A \cup B) = \frac{n(A \cup B)}{n(S)} = \frac{4}{12} = \frac{1}{3}$$

COMPLEMENTO DE UN EVENTO.

TEOREMA 3

Sea A un evento cualquiera y S un espacio muestral, tal que $A \subset S$, si A^c es el complemento del evento A, entonces la probabilidad de A^c es igual a 1 menos la probabilidad de A. Es decir.

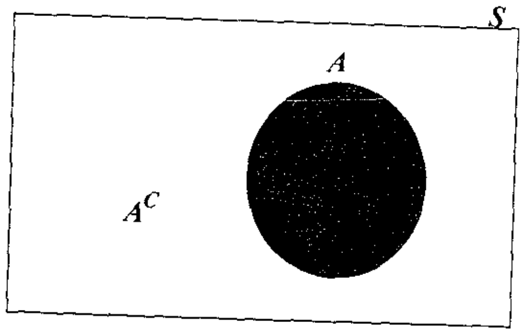
$$P(A^c) = 1 - P(A)$$

demostracion

Se sabe que:

$$A^c = \{x / x \notin A, x \in S\}$$

utilizando un diagrama de Venn.



por la definición del complemento, ningún elemento que pertenezca a A puede estar en A^C*, por lo que* $A \cap A^C = \varnothing$*, son eventos mutuamente exclusivos, además.*

$$A \cup A^c = S$$

aplicando probabilidad.

$$P(A \cup A^c) = P(S)$$

por le axioma 2.

$$P(S) = 1$$

sustituyendo

$$P(A \cup A^c) = 1$$

ahora por le axioma 3.

$$P(A) - P(A^C) = 1$$

despejando a $P(A^C)$

$$P(A^C) = 1 - P(A)$$

EJEMPLO 4.13

Experimento. *Se lanzan 3 monedas.*

Por el principio fundamental del conteo, se sabe que:

$$n(S) = \underset{n_1}{2} \; \underset{n_2}{2} \; \underset{n_3}{2} = 8$$

y cuyos elementos son.

$$S = \{sss, ssa, sas, saa, ass, asa, aas, aaa\}$$

- *Sea A el evento de que al lanzar 3 monedas caigan 2 solos exactamente*

$$A = \{ssa, sas, ass\} \quad , \quad n(A) = 3$$

- *Sea B el evento de que al lanzar 3 monedas caiga por lo menos un sol.*

$$B = \{sss, ssa, sas, saa, ass, asa, aas\} \quad , \quad n(B) = 7$$

- *Sea C el evento de que al lanzar 3 monedas caiga a lo más un sol*

$$C = \{saa, asa, aas, aaa\} \quad , \quad n(C) = 4$$

Se obtendrá:

i) $P(A \cup B)$

ii) $P(A \cup C)$

iii) $P(A^C)$

iv) $P(B^C)$

v) $P(S)$

Respuestas.

i) Como el evento A y el evento B no son mutuamente exclusivos, se aplicará el teorema 1.

$$P(A \cup B) = P(A) - P(B) - P(A \cap B)$$

por lo cual, se considera que

$$P(A) = \frac{3}{8} \qquad , \qquad P(B) = \frac{7}{8}$$

y

$$P(A \cap B) = \frac{3}{8}$$

por lo tanto.

$$P(A \cup B) = \frac{3}{8} + \frac{7}{8} - \frac{3}{8} = \frac{7}{8}$$

ii) Los eventos A y C no tienen elementos comunes, por lo que son eventos mutuamente exclusivos y por lo tanto, se debe aplicar el axioma 3.

$$P(A \cup C) = P(A) + P(C)$$

$$= \frac{n(A)}{n(S)} + \frac{n(C)}{n(S)} = \frac{3}{8} + \frac{4}{8} = \frac{7}{8}$$

iii) Aplicando directamente el teorema 3

$$P(A^c) = 1 - P(A)$$

$$= 1 - \frac{n(A)}{n(S)} = 1 - \frac{3}{8} = \frac{5}{8}$$

iv) Lo mismo que en el caso anterior, utilizando el teorema 3

$$P(B^c) = 1 - P(B)$$

$$= 1 - \frac{n(B)}{n(S)} = 1 - \frac{7}{8} = \frac{1}{8}$$

v) Se tiene por definición de probabilidad clásica, que la probabilidad de cualquier evento, es igual al número del elementos del evento entre el numero de elementos del espacio muestral Por lo tanto, aplicando esta definición.

$$P(S) = \frac{n(S)}{n(S)} = \frac{8}{8} = 1$$

4.2.3 PROBABILIDAD CONDICIONAL Y EVENTOS INDEPENDIENTES.

(Definición).

Sea E un evento arbitrario de un espacio muestral S con $P(E) > 0$. La probabilidad de que un evento A suceda una vez que E ha sucedido o en otras palabras, la probabilidad condicional de A dado E, se define como:

$$P(A/E) = \frac{P(A \cap E)}{P(E)}$$

despejando a $P(A \cap E)$ y haciendo $(A \cap E) = (E \cap A)$ se tiene que

$$P(A \cap E) = P(E \cap A) = P(E)\,P(A/E)$$

Antes de avanzar más en este concepto, es conveniente definir lo que se entiende por eventos independientes.

EVENTOS INDEPENDIENTES.

Se dice que dos eventos A y B son independientes, si la probabilidad de que B suceda no está influenciada porque A haya sucedido o no.

En otras palabras, si A y B son dos eventos cualesquiera, tales que A y $B \subset S$, se dice que el evento A y el evento B son independientes si:

$$P(A \cap B) = P(A)\, P(B)$$

Tomando en cuenta el ejemplo 4.12 se lanza un dado y una moneda, si el dado es lanzado primero y posteriormente la moneda o si se cambia la posición de lanzamiento, no afecta en nada, las formas como puede suceder el experimento. Ahora si se define A como el evento de que al lanzar un dado caiga un número par y sea B el evento de lanzar una moneda caiga sol, el número de elementos del espacio muestral cuando es lanzado un dado y una moneda es:

$$n(S) = \underset{n_1}{\underline{6}} \quad \underset{n_2}{\underline{2}} = 12 \quad \text{por el principio fundamental conteo}$$

y los elementos que cumplen esta condición son

$$2s,\ 4s,\ 6s$$

$$P(\text{un número par con sol}) = \frac{3}{12} = \frac{1}{4}$$

Desde este punto de vista, la probabilidad de obtener un número par y un sol, es $P(A \text{ y } B) = \frac{3}{12}$ *ya que* $P(A) = \frac{3}{6}$ *y* $P(B) = \frac{1}{2}$ *al multiplicar la probabilidad de A por la probabilidad de B resulta.*

$$\frac{3}{6} \times \frac{1}{2} = \frac{3}{12}$$

es decir

$$P(A \cap B) = P(A)\, P(B)$$

cumpliéndose la condición necesaria para que dos eventos sean independientes

EJEMPLO 4.14

Experimento. *Se lanza un par de monedas. ¿Cuál es la probabilidad de que la primera moneda caiga sol y la segunda caiga sol también.*

- *Sea A el evento de que la primera moneda caiga sol.*
- *Sea B el evento de que la segunda moneda caiga sol.*

$(A \cap B)$ sería el evento de suceda A y también suceda B y su probabilidad es.

$$(A \cap B) = \{ss\} \quad ; \quad n(A \cap B) = 1$$

y los elementos del espacio muestral en este caso son

$$S = \{ss, sa, as, aa\} \quad ; \quad n(S) = 4$$

por lo tanto

$$P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{1}{4}$$

¿A y B son independientes?

La probabilidad de que en el primer lanzamiento caiga sol (evento A) es de $\frac{1}{2}$. De igual forma, la probabilidad de que en el segundo lanzamiento caiga también sol (evento B) es de $\frac{1}{2}$. Por lo tanto, la probabilidad de obtener dos soles es $\frac{1}{2} \bullet \frac{1}{2} = \frac{1}{4}$, que al comparar la probabilidad A y B son iguales lo que quiere decir que A y B son independientes.

EJEMPLO 4.15

Se tiene una baraja americana de 52 cartas y se toman dos cartas con reemplazo. Sean.

- *A el evento de que la primera carta sea as.*
- *B el evento de que la segunda carta sea as*

¿Cuál es la probabilidad de obtener dos ases al tomar dos cartas al azar?

> *NOTA IMPORTANTE*
>
> - ***Con Reemplazo.** Significa que el objeto extraído al azar, se coloca nuevamente en el conjunto de oportunidades de donde fue extraído, quedando disponible en otra extracción.*
> - ***Sin Reemplazo.** Significa que el objeto extraído, no se coloca nuevamente en el conjunto de oportunidades de donde fue extraído, no teniendo oportunidad de ser elegida en otra extracción.*

En la primera extracción se pueden tomar 4 ases de 52 cartas, dando una probabilidad de extracción de $\frac{4}{52}=\frac{1}{13}$.

Como la segunda extracción es con reemplazo, un as puede se elegido de entre cuatro ases posibles, pues la carta extraída en el la primera extracción se regresa al conjunto de oportunidades de donde fue extraída. Por lo que la probabilidad de la segunda extracción es de $\frac{4}{52}=\frac{1}{13}$.

La probabilidad de que se tome un as en la segunda extracción después de un as fue tomado en la primera extracción, en otra palabras $P(B|A)$ *, es claramente igual a la probabilidad de* B *. Es decir.*

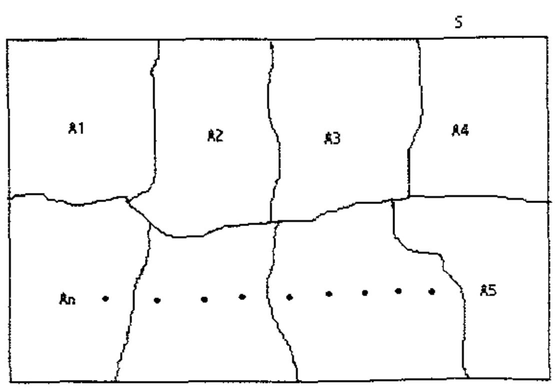
$$P(B|A) = P(B)$$

lo que significa que A y B son independientes.

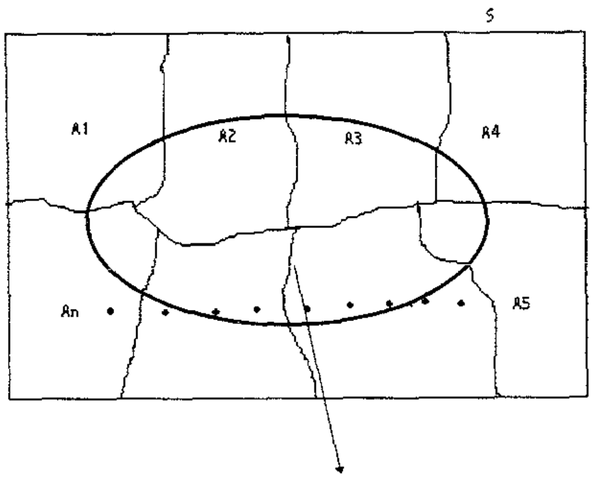
Ahora, sean $A_1, A_2, A_3, \ldots, A_n$ *eventos disyuntos, que forman una partición de S. Esto es* $Ai \cap Aj = \varnothing$, $(\forall i \text{ y } \forall j)$ *para toda i y para toda j.*

y además

$$S = A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5 \cup \ldots \cup An.$$

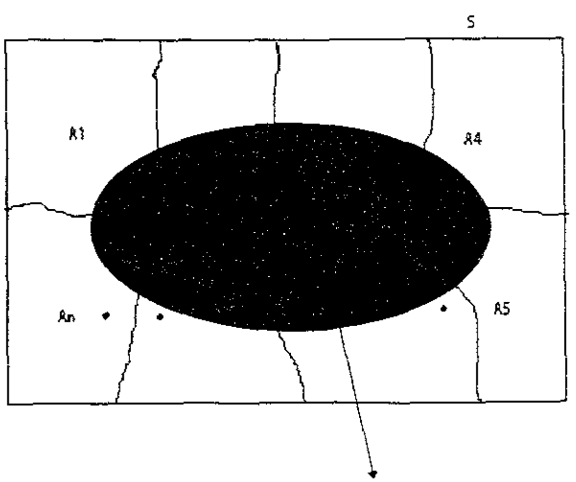


Sea E otro evento, tal que $E \subset S$ *y* $E \bigcap A_1 \neq \phi$



Evento E

por lo tanto $E = S \bigcap E$ *y*



Evento E

$$E = S \cap E = (A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5 \cup \ldots \cup An) \cap E$$

$$= (A_1 \cap E) \cup (A_2 \cap E) \cup (A_3 \cap E) \cup (A_4 \cap E) \cup \ldots \cup (An \cap E)$$

dado que la igualdad entre los eventos se da, al aplicar la función de probabilidad a ambos eventos, se tiene que:

$$P(E) = P(A_1 \cap E) + P(A_2 \cap E) + P(A_3 \cap E) + P(A_4 \cap E) + \ldots + P(An \cap E)$$

ya que todos los eventos $(Ai \cap E)$ *son mutuamente exclusivos y* $Ai \cap E = E \cap Ai$ *entonces* $P(Ai \cap E) = P(E \cap Ai) = P(E/Ai)P(Ai)$. *De tal manera que:*

$$P(E) = P(A_1)P(E|A_1) + P(A_2)P(E|A_2) + P(A_3)P(E|A_3) + \ldots + P(An)P(E|An)$$

EJEMPLO 4. 16

Supóngase un pequeña empresa de tejidos que obtiene su producción con tres máquinas hiladoras M_1, M_2 *y* M_3 *que producen respectivamente 50%, 30% y 20% del número total de artículos producidos. Los porcentajes de productos defectuosos producidos por estas máquinas son 3%, 4% y 5%. Si se selecciona un artículo al azar. ¿Cuál es la probabilidad de que el artículo sea defectuoso?.*

Sea D el evento de un artículo defectuoso. Entonces por la fórmula anterior.

$$P(D) = P(M_1)P(D|M_1) + P(M_2)P(D|M_2) + P(M_3)P(D|M_3)$$

$$= (.50)(.03) + (.30)(04) + (20)(.05)$$

$$= 037$$

EJEMPLO 4.17

En el Colegio de Ciencias y Humanidades, el 4% de los hombres y el 1% de las mujeres miden más de 1.75 metros de estatura. Así mismo, el 55% de los estudiantes son mujeres. Si se selecciona a un estudiante al azar. ¿Cuál es la probabilidad de que mida más de 1.75 metros.

- *Sea X el evento de una persona de más de 1.75*
- *$P(H) = .45$ la probabilidad de se hombre*
- *$P(M) = .55$ probabilidad de ser mujer.*
- *$P(X/H) = .04$ probabilidad de que mida más de 1.75 metros dado que es hombre.*
- *$P(X/M) = .01$ probabilidad de que mida más de 1.75 metros dado que es mujer.*

$$\begin{aligned} P(X) &= P(H)P(X/H) + P(M)P(X/M) \\ &= (.45)(.04) + (.55)(01) \\ &= .018 + .0055 = \mathbf{.0235} \end{aligned}$$

TEOREMA DE BAYES

Supóngase que $A_1, A_2, A_3, \ldots, A_n$ es una partición de un espacio muestral S En cada caso $P(Ai) \neq 0$. La partición es tal que $A_1, A_2, A_3, \ldots, A_n$ son eventos mutuamente exclusivos. Sea E cualquier evento. Entonces para cualquier Ai.

$$P(Ai|E)=\frac{P(Ai)P(E|Ai)}{P(A_1)P(E|A_1)+P(A_2)P(E|A_2)+\ldots+P(An)P(E|An)}$$

EJEMPLO 4.18

Considérese la fábrica del ejemplo 4.15 y supóngase que se selecciona un artículo al azar y resulta ser defectuoso. ¿Cuál sería la probabilidad de que el artículo haya sido producido por la máquina M_1. Ya antes se definió a D como el evento de una artículo defectuoso. Por el teorema de Bayes se tiene.

$$P(M_1/D)=\frac{P(M_1)P(D|M_1)}{P(M_1)P(D|M_1)+P(M_2)P(D|M_2)+P(M_3)P(D|M_3)}$$

$$P(M_1/D)=\frac{(.50)(.03)}{(.50)(.03)+(.30)(.04)+(.20)(.05)}$$

$$=\frac{.015}{.037}=4054$$

4.3 TÉCNICAS DE CONTEO.

Es muy importante entender que en todo problema probabilístico, se de deben encontrar tanto el número de elementos que pertenecen al espacio muestral S, como el número de elementos que determinan a cualquier evento A, B, C, . . , etcétera, para *saber cual es el valor de la probabilidad. En algunos casos, como los que fueron tratados (excepto los que se utilizó el principio fundamental del conteo) esto resulta relativamente sencillo, no así, en donde contar los elementos se torna algo más complicado, haciéndose necesario utilizar técnicas apropiadas para lograr tal propósito.*

Es por tal motivo que se estudiarán técnicas del conteo. también llamadas cálculo combinatorio. En algunos casos, cuando se trata éste tema, sólo se ven permutaciones y combinaciones, sin embargo, la forma en como serán tratadas aquí, es analizando en primera instancia las ordenaciones como una técnica que se basa en el orden y posteriormente las combinaciones, donde llevar un orden de conteo no es posible. Así mismo, se definirá y estudiará el concepto de factorial de un número, cosa necesaria para entender mejor algunas fórmulas que serán usadas al definir dichas técnicas.

FACTORIAL

El factorial de un número entero, se define como, el producto de ese número por todos sus menores enteros y es denotado como $n!$ *donde.*

$$n! = n(n-1)(n-2) \ldots 4 \cdot 3 \cdot 2 \cdot 1$$

por definición, cuando $n = 0$*, este se determina como* $0! = 1$ *.*

EJEMPLO 4.19

$$4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$$

$$3! = 3 \cdot 2 \cdot 1 = 6$$

$$5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$$

Las técnicas de conteo para su estudio se dividen en:

Técnicas de conteo

- *Ordenaciones* (*muy importante el orden*)
  - *Principio fundamental del conteo*
  - *Con repetición*
  - *Sin repetición o permutaciones*
- *Combinaciones* (*no importa el orden*)

La primera técnica es el principio fundamental del conteo, la cual ya fue definida con anterioridad. Sin embargo en este apartado, se efectuarán ejercicios con mayor grado de dificultad y se tratará de que sea comprendida su versatilidad y grandes alcances. Tomando nuevamente la definición.

PRINCIPIO FUNDAMENTAL DEL CONTEO.

Si en un experimento, un primer ensayo puede ocurrir de n_1 maneras diferentes y si continuando el experimento, un segundo ensayo puede ocurrir de n_2 maneras diferentes, y si continuando el experimento un tercer ensayo puede ocurrir de n_3 maneras diferentes, y así sucesivamente, entonces el número de formas como puede ocurrir el experimento es igual a $n_1 + n_2 + n_3 \ldots$

Como ya se mencionó, un ensayo consiste en las repeticiones que se efectuen en un experimento. Así mismo, cuando se maneja el principio, se recomienda utilizar cajas o cuadros para auxiliarse en su cálculo. Por ejemplo.

EJEMPLO 4.20

Experimento. *Se lanzan 6 monedas balanceadas. ¿cuántos resultados de interés existen?*

Como en el lanzamiento de una moneda sólo existen 2 posibles resultados de interés, por el principio fundamental del conteo, cada lanzamiento de una moneda es equivalente a un ensayo, se tiene que el experimento en cuestión puede ocurrir de·

$$\text{No. de formas} = \underset{n_1}{\boxed{2}}\,\underset{n_2}{\boxed{2}}\,\underset{n_3}{\boxed{2}}\,\underset{n_4}{\boxed{2}}\,\underset{n_5}{\boxed{2}}\,\underset{n_6}{\boxed{2}} = 64$$

esto es, al lanzar la primera moneda sólo hay 2 resultados posibles de interés, lo mismo pasa con la segunda, tercera, cuarta, quinta y sexta moneda, en consecuencia, el número de formas en que puede ocurrir el experimento es $2^6 = 64$ *formas o resultados posibles.*

EJEMPLO 4.21

Experimento. *Se lanzan 3 dados.*

En este ejercicio, el primer ensayo n_1, *consiste en lanzar el primer dado,* n_2 *correspondería a lanzar el segundo dado y* n_3 *a lanzar el tercero, quedando así.*

	dado 1	*dado 2*	*dado 3*	
No. de formas =	6	6	6	$= 6^3 = 216$
	n_1	n_2	n_3	

EJEMPLO 4. 22

En un estado de la Repúblıca Mexicana, se pretenden elaborar placas para automóviles que contengan tres letras y cuatro números, se permite que las letras sólo puedan comenzar con A, B o C, y se supone un total de 26 letras en el alfabeto español Otra restricción es que los números no pueden comenzar en cero. L a pregunta que se hacen es: ¿cuántas placas se pueden formar?.

Así mismo, se pretende ımplantar el programa hoy no circula y se tiene la experiencia de asignar cinco colores y la termınación de dos dígitos en los números para el control y la aplicación de esta medıda. Como segunda pregunta que se hacen es ¿cuántas placas se pueden formar por color?

Por últımo, el Sr. Gutiérrez un contribuyente muy cumplido, desea saber que probabılidad tiene su automóvil de descansar los lunes ya que ese día él no trabaja.

Este problema real, se puede resolver fácilmente utılızando el princıpıo fundamental del conteo, comenzando por responder pregunta por pregunta.

- ***¿cuántas placas se pueden formar si se tienen 26 letras y diez dígitos?.***

Tomando en cuenta las restricciones que se tienen para la formación de las placas, por ejemplo el hecho de que las letras tengan que empezar con A, B o C es decir 3 formas diferentes (que además sería el equivalente al primer ensayo n_1) así como, los números no pueden comenzar con cero. Obviamente, para utilizar el principio fundamental del conteo en esta caso, se debe adaptar éste a las condiciones del problema, esto es:

	1^a letra	2^a letra	3^a letra	1^{er} número	$2^{°}$ número	3^{er} número	$4^{°}$ número
No de placas =	3	26	26	9	10	10	10
	n_1	n_2	n_3	n_4	n_5	n_6	n_7

$$= \; 3 \cdot 26 \cdot 26 \cdot 9 \cdot 10 \cdot 10 \quad 10 \; = \; \mathbf{18\,252\,000 \; placas}$$

- ***¿cuántas placas corresponden a cada color?***

Se mencionó con anterioridad, que se toman 5 colores en el para implantar el programa (azul, rosa, rojo, verde y amarillo) y que a cada color se le asignan dos números, esto hace que el número de placas por color sea el mismo en todos los casos.

Color	***números asignados***
amarillo	5,6
rosa	8,7
rojo	4,3
verde	1,2
azul	0,9

efectuando los cálculos.

	1^a letra	2^a letra	3^a letra	1^{er} número	$2^{\cdot}$ número	3^{er} número	$4^{\cdot}$ número
No de placas =	3	26	26	9	10	10	2
	n_1	n_2	n_3	n_4	n_5	n_6	n_7

$$= \ 3 \cdot 26 \quad 26 \cdot 9 \quad 10 \cdot 10 \cdot 2 \ = \ \mathbf{3\,650\,400 \ placas}$$

Por lo tanto, el número de placas para el día lunes o color azul, es de 3 650 400 de un total de 18 252 000. Haciendo esto, que la probabilidad de que el carro del Sr. Gutiérrez descanse el día lunes sea de $\frac{3650400}{18252000} = .2$.

ORDENACIONES CON REPETICIÓN

Esta técnica es comúnmente usada cuando los objetos que se quieran contar pueden ser repetidos, por ejemplo, cuando se trabaja con números, o con letras o cuando se mide el número de soles o de águilas en el lanzamiento de monedas, etcétera, y se define como:

$$nORm = n^m$$

y se lee como, ***las ordenaciones con repetición de n objetos tomados m a la vez.*** *En ocasiones, las ordenaciones con repetición son llamadas permutaciones con repetición.*

EJEMPLO 4.23

Se tienen las letras a, b, c, d y e y se quieren formar "palabras" de dos letras, ¿cuántas "palabras" podrán formarse?.

Es importante subrayar que algunas de las palabras que se construyan no tendrán sentido y que una letra podrá ser repetida. Puede observarse que las "palabras" que se pueden considerar en total son las siguientes.

aa	*ab*	*ac*	*ad*	*ae*
ba	*bb*	*bc*	*bd*	*be*
ca	*cb*	*cc*	*cd*	*ce*
da	*db*	*dc*	*dd*	*de*
ea	*eb*	*ec*	*de*	*ee*

En este caso como sólo se planteó tratar con 5 letras, es posible revisar el número de "palabras" construidas, pero si las letras fueran todas la del alfabeto (26 letras) o alguna otra consideración, ya no sería tan fácil, por lo que se hace necesario considerar como manera más sencilla de determinar el número total de "formas" como "algo" puede ocurrir, así por ejemplo, en el caso que se esta considerando, se tiene que son 5 letras u objetos y se quieren ver "palabras" de 2 dichos elementos, por lo tanto, se tiene que.

$$ {}_5OR_2 = 5^2 = \mathbf{25} $$

EJEMPLO 4.24

Un arquitecto está construyendo una serie de departamentos y tiene la necesidad de numerarlos, él sugiere que éstos tengan cuatro dígitos de los siguientes 1, 2, 3, 4, 5, 6, 7, 8 y 9 , ¿cuántas casas podrá numerar?.

Es claro que el número 1111 tendría sentido toda vez que no está limitada su *posibilidad, lo que indica, que los números pueden ser repetidos y por lo tanto, se puede llegar al resultado con las ordenaciones con repetición de 9 números tomados cuatro a la vez, y queda expresado como.*

$$_9OR_4 = 9^4 = 6\ 561 \text{ números.}$$

ORDENACIONES SIN REPETICIÓN

Esta técnica es usada cuando en el ordenamiento, los elementos no pueden ser repetidos, esto sucede comúnmente con personas , aunque también en muchos otros casos. Las ordenaciones sin repetición son llamadas permutaciones y su fórmula es

$$nOm = \frac{n!}{(n-m)!}$$

y se lee como, ***las ordenaciones sin repetición de n objetos tomados m a la vez.***

EJEMPLO 4.25

Tomando el ejemplo 4.23 pero en el caso de que las letras no puedan ser repetidas, ¿cuántas palabras se podrán formar?.

Se puede fácilmente observar que la palabra "de" no tiene el mismo sentido que " ed" lo que demuestra que se debe llevar orden para contar y como no es posible repetir letra, las palabras aa, bb, cc, dd y ee no podrán ser consideradas, por lo que se obtendrá el resultado por medio de las ordenaciones sin repetición o permutaciones de 5 letras tomadas 2 a la vez.

$$ {}_5O_2 = \frac{5!}{(5-2)!} = \frac{5!}{3!} = \frac{5 \bullet 4 \bullet 3 \bullet 2 \bullet 1}{3 \bullet 2 \bullet 1} = \frac{120}{6} = 20 \quad palabras $$

haciendo una comparación con el ejercicio 4.23, prácticamente serian todos los elementos excepto aquellas en que las letras se repiten, que a continuación se marcan con un circulo y que no pertenecerian al número de formas de las ordenaciones sin repetición de 5 números tomados 2 a la vez, quedando sólo 20.

aa	*ab*	*ac*	*ad*	*ae*
ba	*bb*	*bc*	*bd*	*be*
ca	*cb*	*cc*	*cd*	*ce*
da	*db*	*dc*	*dd*	*de*
ea	*eb*	*ec*	*de*	*ee*

EJEMPLO 4.26

En una cajero automático de un banco, se encuentran formadas en una fila 10 personas, el tiempo de espera de cada persona para poder ser atendida es de 5 minutos aproximadamente. ¿de cuántas maneras se puede formar esa fila?.

Es importante hacer ciertas observaciones en este ejercicio, en primer lugar ¿qué técnica de conteo se debe utilizar en este caso? ¿son ordenaciones? y si son ordenaciones, ¿éstas son con repetición o sin repetición?. Se le cuestionaria al lector, ¿le daría lo mismo estar hasta adelante de la fila o hasta atrás? es decir, ¿le daría lo mismo ser atendido de inmediato que ser atendido después de 3 horas y media? La respuesta debería ser un rotundo no, por lo que importa mucho el orden en como estén formadas las personas, pues en ese orden serán atendidas. Por otra parte, es obvio que una persona por si misma no puede ocupar varios lugares en la fila, por lo que se convierte en una ordenación donde no se vale repetir de 10 personas tomadas todas a la vez , quedando.

$$ {}_{10}O_{10} = \frac{10!}{(10-10)!} = \frac{10!}{0!} = \frac{10!}{1} = 10! = 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = \mathbf{3\,628\,800} $$

EJEMPLO 4.27

Un sindicato desea formar su mesa directiva de entre 200 socios, en la cual habrá un presidente, un secretario y un tesorero. Los comicios, se prevé, serán reñidísimos, pues todos los socios quieren pertenecer a dicha mesa directiva ya que esto, además de darles prestigio, les dará mejores ingresos económicos según sea el puesto dentro de la mesa directiva. ¿Cuántas maneras hay de formarla?

Como todos los socios quieren pertenecer a la mesa directiva, se supone que en primera instancia se disputará el puesto de presidente, el cual será escogido de entre 200 socios, en segunda instancia, se luchara electoralmente por el puesto de secretario, el cual deberá ser elegido de entre 199 socios, puesto que ya deberá haber *sido elegido un presidente. Por último, para el puesto de tesorero sólo podrá ser escogido de entre 198 socios. Hay que tener en cuenta que una sola persona no podrá ocupar los tres puestos a la vez. Este ejercicio resulta sencillo si se aplica la formula de las ordenaciones sin repetición de 200 socios tomados de tres en tres, quedando*

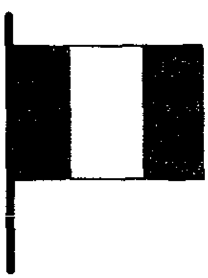
$$ {}_{200}O_3 = \frac{200!}{(200-3)!} = \frac{200!}{197!} = \frac{200 \cdot 199 \cdot 198 \cdot \cancel{197!}}{\cancel{197!}} = $$

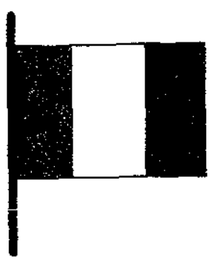
$$ = \; 7\,880\,400 \quad \textbf{formas} $$

EJEMPLO 4.28

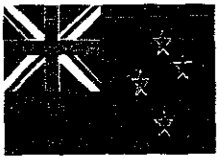
Se tienen los colores blanco, azul, verde, rojo, café y amarillo y se quieren formar banderas de tres colores continuos y verticales. ¿Cuántas banderas se pueden formar? ¿cuántas comienzan con rojo?.

Para resolver este ejercicio, es necesario entender que una bandera tiene sentido si se visualiza puesta en una asta, es decir, una bandera azul, blanca y roja será diferente a una roja, blanca y azul.





además puede haber banderas que tengan estos tres colores acomodados de diferente manera.



sin embargo, se pide colores continuos y verticales.

Tomando en cuenta todo lo anterior y aplicando las ordenaciones con repetición de 6 colores tomados de tres en tres, se tiene.

$${}_6O_3 = \frac{6!}{(6-3)!} = \frac{6!}{3!} = \frac{6 \cdot 5 \cdot 4 \cdot 3!}{3!} = 6 \cdot 5 \cdot 4 = \textbf{120 banderas}$$

¿cuántas comienzan con rojo?.

Es claro que si la bandera principia con rojo, sólo tendrá una forma de empezar y además , si se pone el rojo como primer color, éste no podrá ser repetido, pues si así fuera, la bandera sería de un solo color. Utilizando el principio fundamental del conteo.

$$\text{No. de banderas que comienzan con rojo} = \underset{n_1}{1} \quad \underset{n_2}{5} \quad \underset{n_3}{4} = 20$$

COMBINACIONES

En muchos ocasiones no es posibles llevar un orden de conteo , por el hecho de que esto será imposible, por ejemplo, supóngase una urna con 10 bolas rojas de billar, como las bolas de billar son idénticas, si se tomara una bola al azar con reemplazo, es decir, se toma, se muestra y se vuelve a dejar en la urna, y se repitiera la acción ¿se podría afirmar que la bola de la segunda extracción no es la misma que la de la primera?, no se podría garantizar el hecho de que la bolas seleccionas en ambos casos fueran dos diferentes del conjunto de bolas de donde son tomadas o bien hubiera sido la misma, por lo que el ordenamiento no tiene sentido.

Cuando lo anterior sucede, es necesario utilizar una técnica que no requiera del orden para poder contar el número de casos y esta es llamada ***combinaciones*** *y se define como.*

$$ {}_nC_m = \frac{n!}{m!(n-m)!} $$

se lee como, las combinaciones de n elementos tomados m a la vez.

Quizás las combinaciones como técnica de conteo, sea la herramienta más utilizada para le cálculo de probabilidades y su característica principal es que como se menciono con anterioridad, no se requiere de orden para contar el número de formas en como puede ocurrir un experimento cualquiera.

EJEMPLO 4.29

Se tienen las letras a, b, c y d. ¿Cuántas combinaciones se pueden formar con dos de estas letra?.

Este ejemplo, ayuda a explicar varias cosas, primero recuérdese que en combinaciones ***no importa el orden,*** *por lo que ab es la misma combinación que ba.*

Así mismo, aunque en este caso se pueden ver las letras y se podría llevar un orden, en este ejercicio se supondrá que eso no puede pasar pues sólo será utilizado para ejemplificar.

El total de combinaciones se despliega a continuación.

ab, ac, ad, bc, bd, cd

nótese que.

ba, ca, da, cb, db, dc

es el mismo número de combinaciones que las anteriores ya que ***no importa el orden.*** *Utilizando la fórmula de las combinaciones de 4 letras tomadas 2 a la vez.*

$$ {}_4C_2 = \frac{4!}{2!(4-2)!} = \frac{4!}{2!2!} = \frac{4 \cdot 3 \cdot 2!}{2 \cdot 2!} = \frac{4 \cdot 3}{2} = \frac{12}{2} = 6 $$

que son las 6 combinaciones mencionadas con anterioridad.

EJEMPLO 4.30

Supóngase un urna que contiene 15 bolas de billar, donde 10 son blancas y 5 rojas. Si se toman tres bolas al azar. ¿de cuántas formas se pueden extraer?.

i) tres bolas.

ii) tres bolas blancas.

iii) tres bolas rojas.

iv) dos rojas y una blanca.

v) dos blancas y una roja.

- ***número total de formas en como se pueden tomar tres bolas cualesquiera que estas sean sin importar el color, aplicando la fórmula de las combinaciones de 15 tomadas tres a la vez.***

$${}_{15}C_3 = \frac{15!}{3!(15-3)!} = \frac{15!}{3!12!} = \frac{15 \bullet 14 \bullet 13 \bullet 12!}{3 \bullet 2 \bullet 1 \bullet 12!} = \frac{2730}{6} = \boldsymbol{455}$$

- ***número de formas en como se pueden obtener sólo bolas blancas .***

Como hay de entre las quince bolas diez blancas, se tomarán en cuenta únicamente estas para calcular el número de formas en como se pueden tomar 3 de entre ellas. Obviamente si se llegasen a tomar tres bolas rojas, o dos rojas y una blanca, o dos blancas u una roja, no revestiría interés alguno para el caso de que esta tratando. Por lo que, este evento se puede calcular con la combinaciones de 10 bolas blancas tomadas tres a la vez .

$${}_{10}C_3 = \frac{10!}{3!(10-3)!} = \frac{10!}{3!7!} = \frac{10 \bullet 9 \bullet 8 \bullet 7!}{3 \bullet 2 \bullet 1 \bullet 7!} = \frac{720}{6} = \boldsymbol{120} \textit{ formas de extraer 3 bolas blancas}$$

- ***número de formas de sólo bolas rojas.***

Aquí como en el cálculo anterior, sólo serían de interés, los casos en los cuales se puedan elegir bolas rojas, es decir, las combinaciones de 5 bolas rojas tomadas 3 a la vez .

$${}_{5}C_3 = \frac{5!}{3!(5-3)!} = \frac{5!}{3!2!} = \frac{5 \bullet 4 \bullet 3!}{3! \bullet 2 \bullet 1} = \frac{20}{2} = \boldsymbol{10} \textit{ formas de extraer 3 bolas rojas}$$

- ***número de formas de obtener dos bolas rojas y un blanca.***

Para poder encontrar le número de combinaciones que se tienen en 'e caso, que estarian dadas por la formas en como se pueden tomar 2 bolas rojas por las formas en como se puede obtener una blanca, esto haciendo uso del principio fundamental del principio fundamental del conteo, donde el primer ensayo corresponderia a las formas de extraer dos bolas rojas y el segundo las formas de como puede ser extraia una blanca, estaria dado por:

$${}_5C_2 = \frac{5!}{2!(5-2)!} = \frac{5!}{2!3!} = \frac{5 \bullet 4 \bullet 3!}{2 \bullet 3!} = \frac{20}{2} = \boldsymbol{10} \text{ No de formas de extraer dos bolas rojas}$$

$${}_{10}C_1 = \frac{10!}{1!(10-1)!} = \frac{10!}{9!} = \frac{10 \bullet 9!}{9!} = \boldsymbol{10} \text{ No de formas de extraer una bola blanca}$$

De donde.

$$\text{No. de formas de 2 rojas y una blanca} = \begin{array}{c|c} \boxed{{}_5C_2} & \boxed{{}_{10}C_1} \\ \boldsymbol{n_1} & \boldsymbol{n_2} \end{array} = 10 \quad 10 = \boldsymbol{100}$$

- ***análogamente el número de formas en como se pueden extraer 2 bolas blancas y una roja.***

Dos bolas blancas de diez que hay, se pueden tomar.

$${}_{10}C_2 = \frac{10!}{2!(10-2)!} = \frac{10!}{2 \bullet 8!} = \frac{10 \bullet 9 \bullet 8!}{2 \bullet 8!} = \frac{90}{2} = \boldsymbol{45}$$

una bola roja de 5, se pude escoger de.

$$_5C_1 = 5$$

por lo que el número de formas que hay de sacar dos blancas y una roja, esta dado por.

$$_{10}C_2 \quad _5C_1 \quad = \quad 45 \quad 5 \quad = \quad \mathbf{225}$$

resumiendo.

Extracciones	***No de formas***
3 blancas	*120*
3 rojas	*10*
2 rojas y 1 blanca	*100*
2 blancas y 1 roja	*225*
total	***455***

Como se puede observar en el cuadro anterior, se muestran resumidas todas las opciones que se tienen en el experimento de sacar 3 bolas al azar de 15 que contiene una urna, de las cuales 10 son blancas y 5 rojas. Haciendo un pequeño análisis de lo que esta ocurriendo, el total de formas esta compuesto por las formas en como se pueden tomar 3 bolas blancas, 3 bolas de color rojo, 2 rojas y una blanca y dos blancas y una roja.

EJEMPLO 4.31

El grupo 564 del curso de Estadística y Probabilidad I del Colegio de Ciencias y Humanidades, Plantel Vallejo, quiere formar una comisión de tres personas que los representara en todos los compromisos que pueda tener el grupo. ¿Dé cuántas maneras se puede formar dicha comisión?.

Este ejemplo a diferencia del ejemplo 4.27, en el que se tenia que formar una mesa directiva con importancia en los puestos a ocupar, aquí, no tiene ninguna importancia el peso de los miembros que pertenezcan a la comisión, por lo que, de la forma en como se elijan los tres miembros de la comisión siempre será la misma comisión. haciendo nula la importancia del orden en como se puedan ser elegidos y por tal motivo, el uso de las combinaciones para calcular formas en como se forma dicha comisión. El total de formas esta dado por·

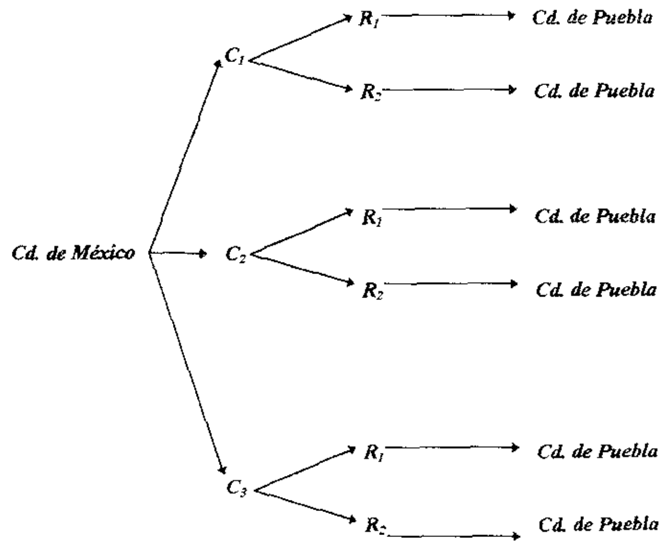
$$ {}_{34}C_3 = \frac{34!}{3!(34-3)!} = \frac{34!}{3!31!} = \frac{34 \bullet 33 \bullet 32 \bullet 31!}{3 \bullet 2 \bullet 1 \bullet 31!} = \frac{35904}{6} = \boldsymbol{5\,984} $$

Es muy importante entender, que el estudio de las técnicas del conteo, se lleva a cabo con el firme propósito de que éstas sirvan como una fuerte herramienta para el cálculo de probabilidades.

En algunos casos, los diagramas de árbol, pueden servir de ayuda para resolver problemas probabilísticos sencillos y proporcionar una manera fácil de visualizar las formas en como puede ocurrir un experimento. Por ejemplo.

EJEMPLO 4.32

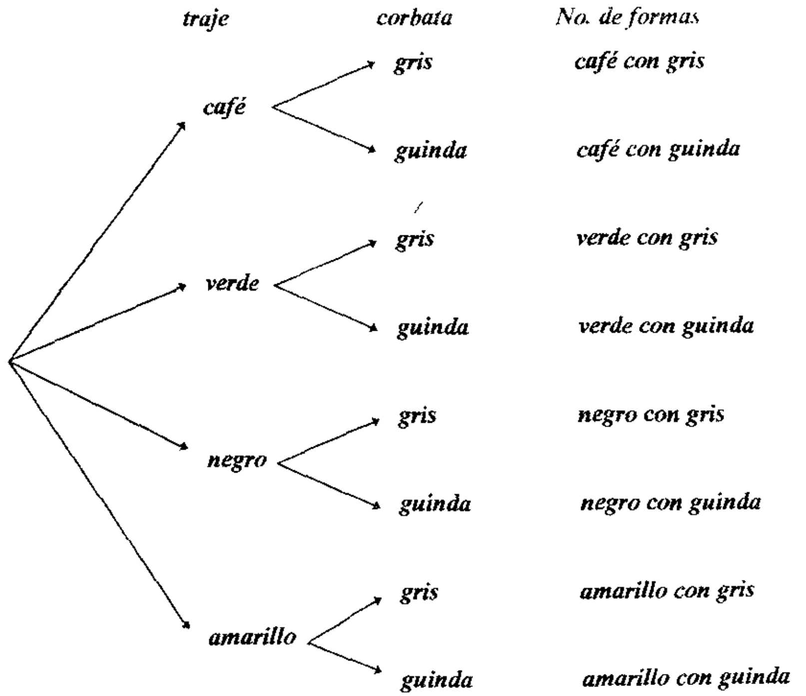
Un agente viajero desea saber cuántas rutas puede tomar para llegar por carretera de la Ciudad de México a la Ciudad de Puebla, él sabe que para lograr su objetivo, necesita pasar por dos ciudades intermedias y sabe por experiencias propia que de la Ciudad de México a la primera ciudad intermedia, hay tres caminos diferentes (C_1 , C_2 y C_3), que puede tomar y de la primera ciudad intermedia a la segunda hay sólo dos rutas a seguir (R_1 y R_2). Sin embargo, para llegar de la segunda ciudad intermedia a la Ciudad de Puebla existe un solo camino (P) ¿Cuál es el número de rutas que el agente puede seguir?.



lo que proporciona 6 rutas diferentes en total

EJEMPLO 4.33

El Sr. Domínguez tiene 4 trajes, (café, verde, negro y amarillo) y dos corbatas (gris y guinda). Él siempre al vestirse primero escoge el traje y posteriormente la corbata. ¿Dé cuántas formas puede el Sr. Dominguez combinar sus trajes con sus corbatas ?



dando como resultado un total de ocho formas.

VARIABLE ALEATORIA

*El estudio de **Variable Aleatoria** a nivel medio superior no es sencillo, muchos autores manejan este concepto con diversos criterios de profundidad que van desde lo más elemental hasta lo más complejo y teórico. En estos apuntes se tratará de cuidar el nivel de enseñanza que requieren los alumnos de la Escuela Nacional Colegio de Ciencias y Humanidades. Aun cuando la forma en como se analizará es ligeramente complicada, se hace con el firme propósito de mostrar el espíritu matemático, sin descuidar los objetivos esenciales de este trabajo. Es decir, tratar de que los conceptos que se manejen aquí sean lo más compresibles a dicho nivel académico Así mismo, los ejemplos que se manejen deberán de ser los más sencillos.*

Es obvio, que existen ejemplos en donde el trabajo que hay que desarrollar y el tiempo que hay que invertir para lograr el resultado deseado, sobrepasa los alcances de esta exposición , dichos ejemplos corresponderían a experimentos conceptuales en donde la característica en estudio o la información que se está utilizando es tanta que se considera de tipo continuo, como podría ser el caso de la estatura, el peso, la edad, entre otras. Por lo cual aquí no serán tratados.

A continuación se revisarán de manera directa las definiciones que permitan ubicarse en el proceso de construcción de un espacio de probabilidades y que posteriormente facilitarán el manejo de la definición de lo que es una variable aleatoria.

- ***Espacio Muestral.*** *Es la colección de todos los posibles resultados de un experimento y se denota con la letra S.*
- ***Espacio de Eventos*** α . *Es una colección que se compone de todos los subconjuntos posibles del espacio muestral asociados con un experimento.*
- ***Función de Probabilidad P.*** *Es una función con dominio en* α *(espacio de*

eventos) y contradominio en el intervalo [0,1] (que es como ya se mencionó con anterioridad, el único intervalo donde está definida la probabilidad), la cual satisface los siguientes axiomas o propiedades

i) $0 \leq P(A) \leq 1$ *La probabilidad de un evento cualquiera A que pertenezca al espacio de eventos es menor o igual a 1 y mayor o igual a cero*

ii) $P(S) = 1$ *La probabilidad del espacio muetral es igual a la unidad (evento certeza).*

iii) Si A_1 A_2, A_3, A_4, *son eventos mutuamente exclusivos o disyuntos, es decir no tienen elementos comunes entre si, y pertenecen todos al es espacio de eventos* α *, entonces,* $P(A_1 U A_2 U A_3 U A_4 \quad) = P(A_1) + P(A_2) + P(A_3) + P(A_4) +$ *la probabilidad de la unión de los eventos es igual a la suma de sus probabilidades.*

y tres teoremas

i) $P(\varnothing) = 0$ *La probabilidad del evento nulo es igual a cero.*

ii) Si A es un evento en que esta en el espacio de eventos α *, entonces.*

$$P(A^C) = 1 - P(A)$$

iii) Para cada dos eventos A, B que estén definidos en el espacio de eventos α

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- **Espacio de Probabilidad.** *Consiste en la tripleta S (espacio muestral),* α *(espacio de eventos) y P (una función conjunto de probabilidad) .*

Se dice que X es una variable aleatoria, si X es una función real definida sobre un espacio muestral S como dominio y contradominio el conjunto de los números reales.

En general se debe considerar que una variable aleatoria X es una función que asigna un cierto valor a los eventos elementales del espacio muestral. La función X debe ser tal que sea posible determinar al conjunto Ar definido por $Ar = \{w \,/\, X(w) \leq r\}$, *como elementos del espacio de eventos* α .

donde:

$r \in \boldsymbol{R}$

w *son los eventos elementales que pertenecen al espacio muestral S .*

EJEMPLO 4.34

Experimento. *Se lanza un dado.*

El espacio muestral S de todos los posibles resultados de interés es representado por:

$\boldsymbol{S} = \{$  $\}$

que por simplicidad es denotado por.

$$S = \{1, 2, 3, 4, 5, 6\}$$

donde la cardinalidad o número de elementos se S está dado por

$$\boldsymbol{n(S) = 6}$$

En este caso la variable que se tratará es la que representa el número de puntos que tiene la cara que esta hacia arriba.

Obsérvese que esta variable aleatoria, lo que esta haciendo es lo siguiente

X(⚀) = *X*(1) = 1

X(⚁) = *X*(2) = 2

X(⚂) = *X*(3) = 3

X(⚃) = *X*(4) = 4

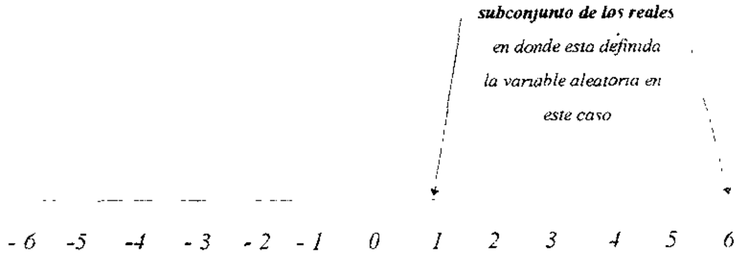
X(⚄) = *X*(5) = 5

X(⚅) = *X*(6) = 6

Notese que una variable aleatoria sólo tiene sentido en los eventos elementales del espacio muestral y no sirve para aplicarse a todos los elementos del espacio de *eventos, pues no tiene sentido querer aplicar la función en cuestión a eventos tales como* $\{1, 6\}$ *ó* $\{2, 4, 6\}$.

Aunque es más o menos fácil definir una variable aleatoria, dado un experimento conceptual, demostrar que una cierta transformación adoptada cumple con la definición de variable aleatoria, es un poco latoso pues es necesario corroborar que para cualquier número real " r " el conjunto Ar correspondiente, es elemento del espacio de eventos α.

En el caso del experimento de lanzar un dado, se tiene lo siguiente



como se observa X es una función que va del espacio muestral S a los reales en símbolos, $X: S \rightarrow \mathbf{R}$. *en este caso, X es una función que va de S a un subconjunto de los números reales (1, 2, 3, 4, 5, 6).*

De acuerdo a la variable aleatoria considerada se puede observar que existen los siguientes conjuntos Ar; a saber para $r < 1$

$$\{ w \,/\, X(w) \leq r \} = \{ \} = \varnothing$$

Si $1 \leq r < 2$ *el Ar en cuestión está dado por*

$$\{w /\ X(w) \leq r\} = \{ ⚀ \} = \{1\}$$

Si $2 \leq r < 3$ *el* Ar *correspondiente será*

$$\{w /\ X(w) \leq r\} = \{ ⚀ , ⚁ \} = \{1, 2\}$$

Si $3 \leq r < 4$ *el* Ar *correspondiente será*

$$\{w /\ X(w) \leq r\} = \{ ⚀ , ⚁ , ⚂ \} = \{1, 2, 3\}$$

Si $4 \leq r < 5$ *el* Ar *correspondiente será*

$$\{w /\ X(w) \leq r\} = \{ ⚀ , ⚁ , ⚂ , ⚃ \} = \{1, 2, 3, 4\}$$

Si $5 \leq r < 6$ *el* Ar *correspondiente será*

$$\{w /\ X(w) \leq r\} = \{ ⚀ , ⚁ , ⚂ , ⚃ , ⚄ \} = \{1, 2, 3, 4, 6\}$$

Si $6 \leq r$ *el* Ar *corresponderá a*

$$\{w /\ X(w) \leq r\} = \{ ⚀ , ⚁ , ⚂ , ⚃ , ⚄ , ⚅ \} = \{1,2,3,4,5,6\}$$

La descripción anterior, parece no ser muy explícita, sin embargo, para entenderla mejor se sugiere revisar el siguiente caso

Considérese los valores 3.1 y 3.8 Para el primer valor, el conjunto $A_{3.1}$, *es el conjunto de todos los eventos elementales, miembros del espacio muestral, tales que al aplicárseles la transformación considerada, esto es la definición de variable aleatoria, el valor que da es menor o igual a 3.1 en este caso, el conjunto en cuestión,* *si se revisa con cuidado está dado por* { ⚀ , ⚁ , ⚂ }, *observe que los elementos* ⚃ , ⚄ *y* ⚅ *no pueden pertenecer a dicho conjunto pues, cuando se les aplica la variable aleatoria en cuestión el valor que se obtiene es mayor que 3.1*

De manera análoga puede observarse exactamente lo mismo con el valor de 3.8, nuevamente el conjunto { ⚀ , ⚁ , ⚂ } *es el conjunto que al aplicarse la transformación, esto es* X(⚀) = 1 , X(⚁) = 2 *y* X(⚂) = 3 , *el resultado correspondiente en cada caso es menor o igual que 3.8 ; y* ⚃ , ⚄ *y* ⚅ *no forman parte de ese conjunto pues* X(⚃) = 4, X(⚄) = 5 *y* X(⚅) = 6 , *da en cada caso valores superiores o mayores a 3.8.*

Si se considera cualquier otro valor del intervalo $3 \leq r < 4$, *el conjunto que se obtendría sería el mismo. De igual manera se puede hacer en los otros intervalos y cerciorarse de que los conjuntos Ar están dados en la forma como ya se especificó.*

La cardinalidad del espacio muestral es 6, esto es $n(S) = 6$, *por lo que se generan 7 regiones en el conjunto de los reales, mismos que dan lugar a 7 conjuntos Ar diferentes, los ya mencionados.*

Como se puede observar los conjuntos Ar obtenidos, todos sin excepción pertenecen al espacio de eventos α , *se sugiere al lector corroborar tal hecho*

Lo anterior permite observar que la definición de variable aleatoria adoptada, determina un conjunto de elementos del espacio de eventos, que no abarca a todos sus componentes, sino a los que están especificados, de alguna manera, por la relación que se establece con ella misma.

El número de elementos que pertenecen al espacio de eventos α *esta dado por.*

$$n(\alpha) = 2^{n(S)}$$

$$= 2^6 = \mathbf{64}$$

otra manera de expresarlo es :

$$n(\alpha) = \{ {}_6C_0 + {}_6C_1 + {}_6C_2 + {}_6C_3 + {}_6C_4 + {}_6C_5 + {}_6C_6 \}$$

$$= 1 + 6 + 15 + 20 + 15 + 6 + 1 = 64$$

que significa que el espacio de eventos está compuesto por los conjuntos siguientes

i) ***el conjunto que tiene cero elementos de S**, esto es el*

$$\varnothing \quad ó \quad \{ \}.$$

es decir el conjunto que no tiene ningún elemento.

i) ***los conjuntos que tienen un sólo elemento de S .***

$$\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$$

iii) ***los que tienen dos elementos diferentes de S.***

$$\{1,3\}, \{1,4\}), \{1,5\}, \{1,6\}, \{2,3\}, \{2,4\}, \{2,5\}, \{2,6\}, \{3,4\}, \{3,5\}, \{3,6\}$$

iv) ***los conjuntos que tienen tres elementos diferentes de S.***

$\{1,2,3\}$, $\{1,2,4\}$, $\{1,2,5\}$, $\{1,2,6\}$, $\{1,3,4\}$, $\{1,3,5\}$, $\{1,3,6\}$, $\{1,4,5\}$, $\{1,4,6\}$, $\{1,5,6\}$, $\{2,3,4\}$, $\{2,4,5\}$, $\{2,3,6\}$, $\{2,4,5\}$, $\{2,4,6\}$, $\{2,5,6\}$, $\{3,4,5\}$, $\{3,4,6\}$, $\{3,5,6\}$, $\{4,5,6\}$

v).- los conjuntos que tienen cuatro elementos diferentes de S.

$\{1,2,3,4\}$, $\{1,2,3,5\}$, $\{1,2,3,6\}$, $\{1,2,4,5\}$, $\{1,2,4,6\}$
$\{1,2,5,6\}$,$\{1,3,4,5\}$,$\{1,3,4,6\}$, $\{1,3,5,6\}$, $\{1,4,5,6\}$, $\{2,3,4,5\}$, $\{2,3,4,6\}$,
$\{2,3,5,6\}$, $\{2,4,5,6\}$, $\{3,4,5,6\}$

vi).- los conjuntos que tienen cinco elementos de S.

$\{1,2,3,4,5\}$, $\{1,2,3,4,6\}$, $\{1,2,3,5,6\}$, $\{1,2,4,5,6\}$, $\{1,3,4,5,6\}$, $\{2,3,4,5,6\}$

vii).- por último el conjunto que tiene todos los elementos de S.

$$\{1, 2, 3, 4, 5, 6\}$$

Con lo anterior se puede ver que al darle valores a r se forma la colección de eventos llamada espacio de eventos y por más grandes que se le pudieran asignar esos valores, seguirá perteneciendo al espacio de eventos, quedando así formados el ***espacio muestral S y el espacio de eventos α.***

$$\begin{aligned}\alpha = \{ &\varnothing, \{1\},\{2\},\{3\},\{4\}, \{5\},\{6\}, \{1,2\}, \{1,3\}, \{1,4\}, \{1,5\}, \{1,6\}, \{2,3\},\\ &\{2,4\}, \{2,5\}, \{2,6\}, \{3,4\}, \{3,5\}, \{3,6\}, \{4,5\}, \{4,6\}, \{5,6\}, \{1,2,3\},\\ &\{1,2,4\}, \{1,2,5\}, \{1,2,6\}, \{1,3,4\}, \{1,3,5\}, \{1,3,6\}, \{1,4,5\}, \{1,4,6\},\\ &\{1,5,6\}, \{2,3,4\}, \{2,4,5\}, \{2,3,6\},\{2,4,5\},\{2,4,6\},\{2,5,6\},\{3,4,5\}, \{3,4,6\},\\ &\{3,5,6\}, \{4,5,6\}, \{1,2,3,4\}, \{1,2,3,5\}, \{1,2,3,6\}, \{1,2,4,5\}, \{1,2,4,6\}\\ &\{1,2,5,6\}, \{1,3,4,5\}, \{1,3,4,6\}, \{1,3,5,6\}, \{1,4,5,6\}, \{2,3,4,5\}, \{2,3,4,6\},\\ &\{2,3,5,6\}, \{2,4,5,6\}, \{3,4,5,6\}, \{1,2,3,4,5\}, \{1,2,3,4,6\}, \{1,2,3,5,6\},\\ &\{1,2,4,5,6\}, \{1,3,4,5,6\}, \{2,3,4,5,6\}, \{1,2,3,4,5,6\}\}.\end{aligned}$$

el espacio de probabilidad asociado consiste en la tripleta

$$(S, P , \alpha)$$

Por lo que faltaría obtener ***la función de probabilidad P*** *que es aquella que asocia a cada punto del espacio muestral , su probabilidad*

Ahora, si **s** *es un punto cualquiera del espacio muestral* ***S*** *y* ***X*** *es una variable aleatoria, entonces $X(s)$ es un valor de la variable aleatoria en S*

En el ejemplo que se esta manejando, el espacio muestral consiste en seis puntos s_1 = ; s_2 = , s_3 = ; s_4 = , s_5 = ; s_6 = . La variable aleatoria puede tomar los valores X_1, X_2, X_3, X_4, X_5, X_6 definida por X_i = número de puntos que tiene la cara que esta hacia arriba en el dado cuando este es lanzado s_i. Lo que puede ser representado en la siguiente tabla.

puntos en S	s_1	s_2	s_3	s_4	s_5	s_6
$X(s)$	1	2	3	4	5	6

como S es un espacio muestral en donde esta definida un función de probabilidad, entonces se puede escribir

$P(X = 1)$ es decir $P[\{ s / X(s) = 1 \}] = P(s_1) = 1\ 6$

$P(X = 2)$ es decir $P[\{ s / X(s) = 2 \}] = P(s_2) = 1\ 6$

$P(X = 3)$ es decir $P[\{ s / X(s) = 3 \}] = P(s_3) = 1\ 6$

$P(X = 4)$ es decir $P[\{ s / X(s) = 4 \}] = P(s_4) = 1\ 6$

$P(X = 5)$ es decir $P[\{ s / X(s) = 5 \}] = P(s_5) = 1/6$

$P(X = 6)$ es decir $P[\{ s / X(s) = 6 \}] = P(s_6) = 1/6$

otra manera de expresarlo es:

$$f(x) = \begin{cases} 1/6 & si\ x = 1,2,3,4,5,6 \\ & en\ otro\ caso \end{cases}$$

o bien

X	$f(X)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

Una variable aleatoria X se llama discreta si sólo toma un numero finito, numerable de valores de X, suponiendo que X toma sólo los valores $X_1, X_2, \ldots X_n, \ldots$ Y con probabilidades $f(X_1), f(X_2), \ldots f(X_n), \ldots$ y suponiendo que A es un subconjunto de los puntos $X_1, X_2, \ldots X_n, \ldots$ Entonces la probabilidad del evento A que se escribe $P(A)$ se define como:

$$P(A) = \Sigma f(X)$$

donde.

$\Sigma f(X)$ = *la suma de todos los valores* X_i, *que estan en el evento A.*

por ejemplo $P(X = 1)$ *significa la probabilidad de que X variable aleatoria tome el valor de 1,* $P(X = 2)$ *la probabilidad de que X variable aleatoria tome el valor de 2 y así sucesivamente.*

$f(X)$ *puede ser llamada función de densidad discreta o simplemente densidad, si cumple las siguientes propiedades.*

$f(X_i) \geq 0$ $\quad i = 1, 2,$ $\quad$ *toda probabilidad debe ser mayor o igual a 0, y*

$\Sigma f(X_i) = 1$ *la suma de todas las probabilidades debe ser igual a 1.*

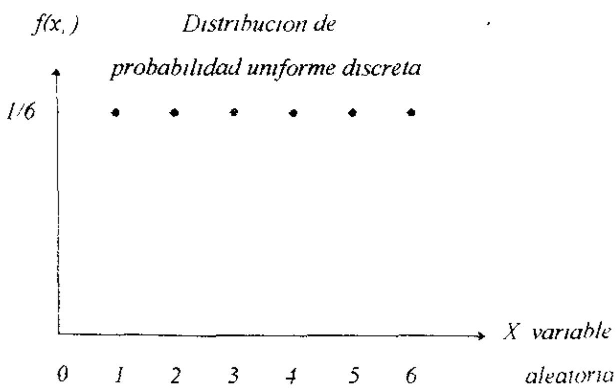
Siguiendo con el ejemplo del dado.

X	f(X)
1	1 6
2	1 6
3	1 6
4	1 6
5	1 6
6	1 6

$\Sigma f(X_i)$ - 1

Es claro que si se toman los elementos del espacio muestral, la probabilidad de cada uno de ellos será igual a $\frac{1}{6}$ *y la suma de las probabilidades es igual a 1. Cuando esto sucede, es decir cuando la variable aleatoria en cuestion, proporciona probabilidades iguales en cada uno de sus elementos , se dice que se estan distribuyendo uniformemente dichas probabilidades y por lo tanto es llamada* ***distribución de probabilidad uniforme*** *y también será* ***discreta.***

Graficando.



Otro ejemplo que se pude desarrollar es el que corresponde al lanzamiento de una moneda legal y en el cual se tratara de llevar a cabo el mismo proceso que en el ejercicio anterior.

EJEMPLO 4.35

Experimento. *Se lanzan una moneda.*

El espacio muestral S de todos los posibles resultados de interés, es representado por

$$S = \left\{ \text{sol} \;,\; \text{águila} \right\}$$

Que por comodidad se usualmente se pone.

$$S = \{ s \;,\; a \}$$

donde el número de elementos que pertenecen al espacio muestral esta expresado por

$$n(S) = 2$$

Si se define la variable aleatoria como aquella que representa el numero de soles que caen al lanzar una moneda, es fácil observar que·

$$X(\text{sol}) = X(1) = \text{sol}$$

$$X(\text{águila}) = X(0) - \text{águila}$$

de acuerdo con la variable aleatoria definida, existen los siguientes conjuntos Ar. Para r . -1.

$$A_{-1} = \{ w \,/\, X(w) \le -1 \} = \{ \} = \varnothing$$

es decir el evento tiene cero elementos de S.

Para r = 0 ,

$$A_0 = \{ w \,/\, X(w) \leq 0 \} = \{ a \} = 0$$

por ultimo para cuando r = 1

$$A_1 = \{ w \,/\, X(w) \leq 1 \} = \{ a , s \} = 0 , 1$$

con lo que se obtienen todos los elementos que pertenecen al espacio muestral se sabe que :

$$n(\alpha) = 2^{n(S)} = 2^2 = 4$$

de donde

$$\alpha = \{ \varnothing, \{ a \}, \{ s \}, \{ s,a \} \}$$

la función de probabilidad P es aquella que asocia a cada punto del espacio muestral S su probabilidad. El espacio muestral en este caso, tiene dos puntos s_1 = águila , s_2 = sol . La variable aleatoria puede tomar los valores $X_1 = 0$ y $X_2 = 1$ ya que ésta esta definida como el número de soles que caen al lanzar una moneda. Como S es un espacio muestral en donde esta definida un función de probabilidad, se puede escribir.

$$P(X = 0) \text{ es decir } P [\{ s \,/\, X(s) = 0 \}] = P(s_1) = 1\ 2$$
$$P(X = 1) \text{ es decir } P [\{ s \,/\, X(s) = 1 \}] = P(s_2) = 1\ 2$$

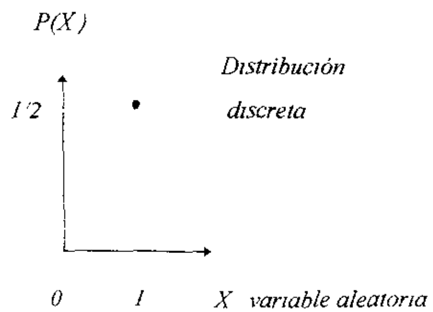
que puede ser representada como:

$$f(X) = \begin{cases} 1/2 & \text{si } X = 0, 1 \\ 0 & \text{en otro caso} \end{cases}$$

o bien

X_i	$f(X_i)$
0	1/2
1	1/2

es sencillo comprobar que la suma de las probabilidades es igual a 1 y que sus valores son enteros numerables por lo que es una distribución discreta, graficando.



EJEMPLO 4.36

Experimento. *Se lanzan dos monedas legales*

El espacio muestral S de todos los todos los posibles resultados de interes en el caso de lanzar un par de monedas, es el siguiente.

$$S = \{ (sol\ sol),\ (sol\ águila),\ (aguila\ sol),\ (aguila\ aguila) \}$$

que es representado como,

$$S = \{ ss, sa, as, aa \}$$

de donde

$$n(S) = 4$$

Si se define la variable aleatoria como el número de soles que pueden caer al lanzan dos monedas legales, entonces

$$X(sol\ sol) = X(2) = sol\ sol$$

$$X(sol\ aguila) = X(1) = sol\ águila$$

$$X(águila\ sol) = X(1) = águila\ sol$$

$$X(aguila\ aguila) = X(0) - águila\ águila$$

lo que indica que la variable aleatoria así definida puede tomar sólo los valores 0, 1 y 2 . Los conjuntos Ar que se pueden obtener con respecto a esta variable son

Para $r < -1$,

$$A_{-1} = \{ w \,/\, X(w) \leq -1 \} = \{ \} = \varnothing$$

nótese que al igual que en ejercicio anterior, ningún número real menor que cero pertenece a la variable por lo cual se formaría el conjunto sin elementos o conjunto vacío.

Para $r = 0$,

$$A_0 = \{ w \,/\, X(w) \leq 0 \} = \{ aa \} = 0$$

para cuando $r = 1$.

$$A_1 = \{ w \,/\, X(w) \leq 1 \} = \{ sa, as \} = 0, 1$$

para $r = 2$

$$A_2 = \{ w \,/\, X(w) \leq 2 \} = \{ sa, as, ss \} = 0, 1, 2$$

con lo que se obtienen todos los elementos que pertenecen al espacio muestral.
se sabe que :

$$n(\alpha) = 2^{n(S)} = 2^4 = 16$$

de donde

$$
\begin{aligned}
\alpha = \{ & \varnothing,\ \{aa\},\ \{as\},\ \{sa\},\ \{ss\}, \\
& \{(aa),(as)\},\ \{(aa),(as)\},\ \{(aa),(ss)\} \\
& \{(as),(sa)\},\ \{(ss),(as)\},\ \{(ss),(sa)\} \\
& \{(aa),(as),(sa)\},\ \{(sa),\ (as),\ (ss)\} \\
& \{(aa),(sa),(ss)\},\ \{(aa),\ (as),(ss)\} \\
& \{(aa),\ (as),\ (sa),\ (ss)\}\}
\end{aligned}
$$

en este caso, P será la función de probabilidad que asocia a cada punto del espacio muestral S probabilidades de $^1/_4$. *El espacio muestral tendrá cuatro puntos de interes*

$s_1 =$ *aguila aguila* ; $s_2 =$ *aguila sol* ; $s_3 =$ *sol aguila* ; $s_4 =$ *sol sol* .

La variable aleatoria puede tomar los valores $X_1 = 0$, $X_2 = 1$ *y* X_3 *de acuerdo en como ésta fue definida.*

La función de probabilidad, se puede escribir.

$$
\begin{aligned}
&P(X = 0) \text{ es decir } P[\{s \mid X(s) = 0\}] = P(s_1) = 1/4 \\
&P(X = 1) \text{ es decir } P[\{s \mid X(s) = 1\}] = P(s_2) = 1/4 \\
&P(X = 1) \text{ es decir } P[\{s \mid X(s) = 1\}] = P(s_3) = 1/4 \\
&P(X = 2) \text{ es decir } P[\{s \mid X(s) = 2\}] = P(s_3) = 1/4
\end{aligned}
$$

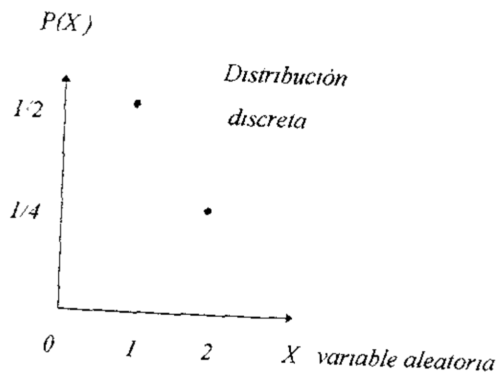
que puede ser representada como:

$$
f(X) = \begin{cases} ^1/_4 & \text{si } X = 0,\ 2 \\ ^1/_2 & \text{si } X = 1 \\ 0 & \text{en otro caso} \end{cases}
$$

o bien

X_i	$f(X_i)$
0	1/4
1	1/2
2	1/4

graficando.



Como un ejercicio adicional, se sugiere al lector, que analice el caso relacionado con el mismo experimento conceptual en donde ahora considerará como variable aleatoria a la relación que a todo resultado posible del espacio muestral, le asigna el valor que consiste en que al número de puntos de la cara superior, le resta seis unidades y al resultado le aplica la función valor absoluto, esto es, en símbolos:

$$Y(w) = | w - 6 | \quad con\ w \in S$$

4.4.1 ESPERANZA MATEMÁTICA

Ahora si X es una variable con distribución anterior, entonces la media o esperanza o valor esperado de X, denotado por E(X) o μ(X) o simplemente E o μ, se define como.

$$E(X) = X_1\ P(X_1) - X_2\ P(X_2) - \qquad - X_n\ P(X_n)$$

$$= \sum_{i=1}^{n} X_iP(X_i)$$

Esto es, E(X) es el promedio ponderado de los valores posibles de X, cada valor ponderado por su probabilidad. En otras palabras la esperanza o valor esperado no es más que el valor que resultará si se repite un número muy grande pero muy grande de veces un cierto experimente que se desee realizar, se suman los resultados obtenidos en cada caso y se divide el total por el número de veces que el experimento se ejecutó

A continuación se calculará el valor esperado para los tres ejemplos anteriores para los que fue definida una variable aleatoria.

En el caso dado.

$$E(X) = (1)(1/6) - (2)(1\ 6) - (3)(1\ 6) - (4)(1\ 6) - (5)(1\ 6) - (6)(1\ 6)$$

$$= 1\ 6 - 2\ 6 - 3\ 6 - 4\ 6 - 5\ 6 - 6\ 6$$

$$= 21\ 6 = 3.5$$

Otro ejemplo que se desarrolló fue el correspondiente al lanzamiento de una moneda legal y en el cual se tratara de llevar a cabo el mismo proceso que en el ejercicio anterior Calculando su valor, se tiene.

$$E(X) = (0 \quad 1\ 2) - (1 \quad 1\ 2)$$
$$= 1\ 2$$

Así mismo, para el experimento cuando son arrojadas dos monedas legales se tiene

$$E(X) = (0 \quad 1\ 4) - (1 \quad 2\ 4) - (2 \quad 1\ 4)$$
$$= 0 - 2\ 4 - 2\ 4$$
$$= 4\ 4 = 1$$

Un ejemplo un poco más complicado al cual se le puede calcular el valor esperado es el siguiente

EJEMPLO 4.37

Experimento. *Se lanzan un par de dados legales.*

Si se define la variable aleatoria como aquella que representa la suma de los puntos que caen al lanzar un par de dados, entonces, el espacio muestral S consta de 36 parejas ordenadas .

$$S = \begin{matrix} (1,1), (2,1), (3,1), (4,1), (5,1), (6,1) \\ (1,2), (2,2), (3,2), (4,2), (5,2), (6,2) \\ (1,3), (2,3), (3,3), (4,3), (5,3), (6,3) \\ (1,4), (2,4), (3,4), (4,4), (5,4), (6,4) \\ (1,5), (2,5), (3,5), (4,5), (5,5), (6,5) \\ (1,6), (2,6), (3,6), (4,6), (5,6), (6,6) \end{matrix}$$

se puede observar que la variable aleatoria puede tomar los valores de 2 y hasta 12 de acuerdo en como fue definida la variable aleatoria, se tiene que

X_i	formas en como puede ocurrir	$P(X_i)$
2	(1,1)	1/36
3	(1,2),(2,1)	2/36
4	(1,3),(2,2),(3,1)	3/36
5	(1,4),(2,3),(3,2),(4,1)	4/36
6	(1,5),(2,4),(3,3),(4,2),(5,1)	5/36
7	(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)	6/36
8	(2,6),(3,5),(4,4),(5,3),(6,2)	5/36
9	(3,6),(4,5),(5,4),(6,3)	4/36
10	(4,6),(5,5),(6,4)	3/36
11	(5,6),(6,5)	2/36
12	(6,6)	1/36

ahora calculando el valor esperado, se tiene.

$$E(X)=2(1/36)+3(2/36)+4(3/36)+5(4/36)+6(5/36)+7(6/36)+8(5/36)+9(4/36)+10(3/36)+11(2/36)+12(1/36)$$

$$=2/36+6/36+12/36+20/36+30/36+42/36+40/36+36/36+30/36+22/36+12/36$$

$$=252/36=7$$

otra manera de obtener el valor esperado, es por medio de una tabla, como la que sigue.

X_i	$P(X_i)$	$X_iP(X_i)$
2	1/36	2/36
3	2/36	6/36
4	3/36	12/36
5	4/36	20/36
6	5/36	30/36
7	6/36	42/36
8	5/36	40/36
9	4/36	36/36
10	3/36	30/36
11	2/36	22/36
12	1/36	12/36

$$\sum_{i=1}^{n} X_iP(X_i) =$$

$$252/36 = 7$$

4.4.2 VARIANZA.

4.4.3 DESVIACIÓN ESTÁNDAR.

Otro parámetro que es considerado cuando se esta trabajando o utilizando el concepto de variable aleatoria, es el de ***varianza*** *denotado por* σ^2 *, o bien su raíz cuadrada llamada* ***desviación estándar*** σ^2 *, (que será estudiada posteriormente)*

La varianza y la desviación estándar son medidas de dispersión, es decir miden la dispersión que tienen los valores de una variable aleatoria con respecto al valor esperado. Este parámetro puede ser obtenido para cuando la variable es discreta y para cuando es continua Sin embargo, en este trabajo sólo se llevará a cabo el caso discreto. Se define como:

$$\sigma^2 = \sum_{i=1}^{n} (Xi - E(Xi))^2 fXi)$$

cuyo desarrollo es.

$$\sigma^2 = (X_1 - E(X))^2 P(X_1) - (X_2 - E(X))^2 P(X_2) - \quad - (X_n - E(X))^2 P(X_n)$$

obteniendo la varianza para cuando es lanzado un dado legal y la va able aleatoria es definida como el número de puntos que aparecen en la cara que cae hacia arriba, se tiene que, el valor esperado que resulto en este caso fue de 3 5 , por lo tanto calculando la varianza .

$$\sigma^2 - (1 - 3.5)^2 (1/6) + (2 - 3.5)^2 (1/6) - (3 - 3.5)^2 (1\ 6) - (4 - 3.5)^2 (1\ 6)$$
$$(5 - 3.5)^2 (1/6) - (6 - 3.5)^2 (1/6)$$

$$= (- 2\ 5)^2 (1/6) - (- 1.5)^2 (1/6) - (- 0.5)^2 (1/6) - (0\ 5)^2 (1\ 6) -$$
$$(1.5)^2 (1/6) - (2.5)^2 (1/6)$$

$$- (6.25) (1/6) - (2.25) (1\ 6) - (0.25) (1\ 6) \cdot (0\ 25) (1\ 6) -$$
$$(2.25) (1\ 6) - (6.25) (1/6)$$

$$- (1.0416) - (0.3749) - (0.0416) - (0.0416) - (0\ 3749) \quad (1\ 0416)$$

$= 2.9162$

Como se mencionó, la desviación estándar es la raiz cuadrada de la varianza, en consecuencia, ésta puede se definida como:

$$\sigma = \sqrt{\sum (X_i - E(X))^2 P(X_i)}$$

o bien

$$\sigma = \sqrt{\sigma^2}$$

$$= \sqrt{varianza}$$

de donde

$$\sigma = \sqrt{2.9162}$$

$$= 1.7076$$

Calculando la varianza para cuando es lanzada una moneda y la vaiable aleatoria es la que representa le número de soles que caen al lanzar una moneda. Se obtuvo como valor esperado 1/2, por lo que :

$$\sigma^2 = \sum (X_i - E(X))^2 P(X_i)$$

$$= (0 - {}^1/_2)^2 \cdot {}^1/_2 \qquad (1 - {}^1/_2)^2$$

$$= {}^1/_4$$

y cuya desviación estándar estaría dada por :

$$\sigma = \sqrt{Var(X)}$$

$$\sigma = \sqrt{1/4}$$

$$\sigma = {}^1/_2$$

y para cuando se lanzan dos monedas y X = número de soles cuando son lanzadas dos monedas legales.

$$\sigma^2 = (0 - 1)^2 (1/4) + (1 - 1)^2 (1/2) + (2 - 1)^2 (1/4)$$
$$= (-1)^2 (1/4) + (0)^2 (1/2) + (-1)^2 (1/4)$$
$$= (1)(1/4) + (0)(1/2) + (1)(1/4)$$
$$= (1/4) + (1/4)$$
$$= 1/2$$

de donde.

5.- FUNCIÓN DE DISTRIBUCIÓN.

Anteriormente ya se bosquejó cual es le proceso de construcción de una función de distribución acumulativa, a partir de la variable aleatoria que se le asocia, y también se vió que la función de distribución se puede determinar a partir de conocer la función de distribución acumulativa o viceversa

Aun cuando el número de funciones o distribuciones que se han desarrollado es bastante importante y su utilidad ha sido amplia, los apuntes para un curso introductorio de estadística y probabilidad no pueden pretender más que revisar los casos de distribuciones relativamente más sencillos o más populares, tal es el caso de las distribuciones uniforme, Bernoulli y binomial, todas ellas discretas y en el caso continuo sólo se revisará la distribución gausiana o normal, la más importante distribución de la estadística.

5.1 TIPO DE DISTRIBUCIÓN DE PROBABILIDAD.

5.1.1 UNIFORME

Una distribución uniforme es aquella en la que, en una variable aleatoria, los puntos en donde $f(X) \neq 0$, $f(X)$ tendrá siempre el mismo valor, esto es, la misma probabilidad. Un ejemplo de este tipo de distribución, es el lanzamiento de un dado legal en donde la variable aleatoria se puede definir como el número de puntos que aparecen en la cara que cae hacia arriba (que ya fue tratado en el capítulo anterior) Es posible observar fácilmente que para cada punto que pertenece a la variable aleatoria así definida, la probabilidad es de 1 6

Xi	*1*	*2*	*3*	*4*	*5*	*6*
P(Xi)	*1 6*	*1 6*	*1 6*	*1 6*	*1 6*	*1 6*

Otro ejemplo pued... r aquel que representa el sexo de un recién nacido ya que de cualquier forma en como sea defina una variable aleatoria en este caso, la probabilidad de hombre es igual a la probabilidad de mujer y esta es de ½

5.1.2 BERNOULLI.

El ensayo de un experimento se llama Bernoulli, cuando en su ejecucion, solo hay dos resultados posible, éxito o fracaso, niño o niña, cierto o falso, vida o muerte, entre otros. Se acostumbra denotar las probabilidades asociadas a dichos resultados por p ***(probabilidad de éxito)*** *y* $q = 1 - p$ ***(probabilidad de fracaso).*** *Es evidente que* p *y* q *no podrán ser negativas y que además* $p + q = 1$ *Cuando el experimento se observa una sola vez y tiene asociada a él una variable aleatoria que se distribuye como una Bernoulli, se dice que ha ocurrido un ensayo Bernoulli*

Por ejemplo, supóngase que se lanza una moneda legal y se observa la cara que aparece. La moneda sólo tendrá dos elementos de interés en este caso, sol o aguila Supóngase además que se desea que un sol aparezca al lanzar la moneda, ... decir que si la moneda cae sol, se apegará a lo deseado y por lo tanto será llamado exito y se denotará por p *. Si por el contrario cayera una águila, lo cual no cumple con lo deseado, será considerado fracaso y se denotará como* $q = 1 - p$.

5.1.3 BINOMIAL

Es conveniente mencionar que la distribución Bernoulli proporciona las bases para entender lo que es la distribución binomial , ya que n ensayos Bernoulli forman una binomial. En otras palabras una Bernoulli puede ser considera... como una binomial en donde se ejecuta un solo ensayo.

La distribución binomial ocurre si se esta interesado en el número de veces que sucede un evento A en n ejecuciones independientes de un experimento aleatorio, con *dos resultados, donde a uno ellos se le denota con " p " y se le define como la probabilidad de éxito y al otro con " q = 1 - p " probabilidad de fracaso (dichas probabilidades deberán se calculadas tomando en cuenta un sólo ensayo binomial)*

Supóngase que X es una variable aleatoria que toma valores 0,1,2, Donde el número de veces que es realizado el mismo experimento esta dado por n (Notese que la variable aleatoria podrá ser el número de soles o de águilas en n lanzamientos de una moneda, o el número de niñas que en n nacimientos , o el numero de verdades en un text verdadero - falso; etcétera y que además la probabilidad de éxito p podrá ser definida de acuerdo al interés que se tenga en alguno de los dos resultados.) Si p es la probabilidad de éxito y q = 1 - p probabilidad de fracaso, se dice que X tiene una distribución binomial si.

$$P(X) = {}_nC_X \; p^X \; q^{(n-X)}$$

*El calculo de la esperanza o valor esperado, la varianza y la desviación estándar, en el caso de la binomial resulta sencillo, pues basta multiplicar el numero de veces en que se repite el experimento, que fue represento con " **n** " por la probabilidad de éxito " **p** ". Es decir,*

$$E(X) = np$$

*por lo que se refiere a la varianza σ^2 es suficiente multiplicar el número de veces en que es repetido el experimento por la probabilidad de éxito por la probabilidad de fracaso " **q** ".*

$$\text{varianza } \sigma^2 = npq$$

por último la desviación estándar σ, es igual a la raíz cuadrada de la varianza, por lo tanto

$$\text{desviación estándar } \sigma = \sqrt{npq}$$

5.1.4 POISSON.

La distribución Poisson, es llamada así, debido a que S. D. Poisson la introdujo en 1837. Esta distribución como un distribución infinita contable, se presenta en muchos experimentos o fenómenos reales, tales como, el número de llamadas telefónicas que llegan por minuto en un tablero de distribución, el número de erratas por página en un libro o en un documento grande, etcétera.

La Poisson se define como:

$$P(X) = \frac{\lambda^{X} e^{-\lambda}}{X!} \qquad X = 1,2,3,$$

donde:

$\lambda > 0$	*una constante*
e	*la función exponencial y su valor es 2.7182*
X	*variable aleatoria*

de igual forma que en la binomial, el cálculo de algunas de sus propiedades es sencillo y aparecen a continuación

$$E(X) = \lambda$$

$$\text{varianza } \sigma^2 = \lambda$$

$$\text{desviación estándar } \sigma = \lambda$$

Esta distribución es una aproximación conveniente de la distribución binomial en casos en donde sucede un gran número de ensayos n y una probabilidad de éxito p pequeña. Por ejemplo

EJEMPLO 5.1

El 2% de los artículos producidos por una fabrica son defectuosos. Encontrar la probabilidad de que en una muestra de 100 artículos elegidos al azar haya 3 defectuosos.

Si se supone el problema binomial, n = 100 , en este caso la probabilidad p de éxito, sería la probabilidad de artículos defectuoso (es obvio que nunca será éxito encontrar un artículo defectuoso, sin embargo para el planteamiento del problema es conveniente escogerlo así) resultando p = 0.02 Puesto que p es pequeña y n es grande, la Poisson da una excelente aproximación donde.

$$\lambda = np = (100)(0.02) = 2$$

ahora aplicando la definición de Poisson, se tiene.

$$P(X = 2) = \frac{2^3 e^{-2}}{3!} = \frac{8(0.135)}{6} = \mathbf{0.180}$$

5.1.5 NORMAL.

La distribución normal o de Gauss, es la distribucion de probabilidad mas importante por muchas razones que se enumerarán mas adelante. Aun en este apartado se mencionaran algunas caracteristicas de esta distribución, se estudiara con más detenimiento posteriormente.

Es conveniente entender que muchas variables aleatorias se encuentran distribuidas normalmente y en las que no sucede esto, se les suele aproximar (bajo ciertas condiciones) a una distribucion normal. Otra razón por la cual es usada la normal, es porque cuando en las distribuciones discretas los datos son grandes o muy grandes y por ende se complica su uso, la distribución normal da un excelente aproximación y esto, minimiza de manera sustancial el trabajo.

La distribución normal de una variable aleatoria X, que toma valores de entre --∞ a ∞ se define como:

$$P(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{(X-\mu)^2}{\sigma^2})}$$

y se dice que X se distribuye como una normal con media μ *y varianza* σ^2 *o en otras palabras* $X \sim N(\mu, \sigma^2)$. *Si la normal es graficada, toma la forma curva acampanada de ahí que también sea llamada campana de Gauss, cuando es aplicada la formula* $Z = \frac{Xi - \mu}{\sigma}$ *, tiene una distribución normal estandarizada con media* $\mu = 0$ *y varianza* $\sigma^2 = 1$ *y toma la siguiente forma.*

$$f(X) = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}x^2}$$

En el apartado 5.3 que esta más adelante se ven mas conceptos de la distribución normal.

5.2 ESTUDIO DE LA DISTRIBUCIÓN BINOMIAL COMO EJEMPLO DE DISTRIBUCIÓN DE PROBABILIDAD DISCRETA.

Cálculo, Graficación e Interpretación

Ya con anterioridad se definió la binomial como una distribución de probabilidad discreta y se mencionaron algunas de sus propiedades que serán utilizadas en este apartado

El objetivo principal de este subcapítulo es el de que por medio de ejemplos, el estudiante a nivel medio superior, maneje la binomial como una herramienta muy importante para el cálculo de probabilidades, así como, mostrar la facilidad con la que por medio de ésta, se llega al resultado deseado

EJEMPLO 5.2

Experimento. *Se lanzan un par de monedas.*

Obtener la probabilidad de caigan dos soles exactamente. La variable aleatoria en cuestión, sería aquella que representa el número de soles que caen al lanzar un par de monedas, es decir $X = 0, 1, 2$ *. Por lo que en este caso se pediría* $P(X = 2)$ *Aplicando la definición de binomial se tiene.*

$$P(X) = {}_nC_X \; p^X \; q^{(n-X)}$$

donde:

nCm *combinaciones de n elementos tomados m a la vez*

p *probabilidad de éxito*

$q = 1 - p$ *probabilidad de fracaso*

X *variable aleatoria.*

Por lo que, en este caso, como n es el número de ensayo realizados (es decir el lanza miento de dos monedas), n = 2. Así mismo, como se está pidiendo el numero de soles que caen , sería conveniente que la probabilidad de éxito se definiera esta como la probabilidad de sol en un sólo ensayo binomial, o sea en el lanzamiento de una sola moneda, por lo que $p = 1/2$ *y* q *probabilidad de que no caiga un sol o probabilidad de fracaso, sería igual a* $1/2$.

De donde.

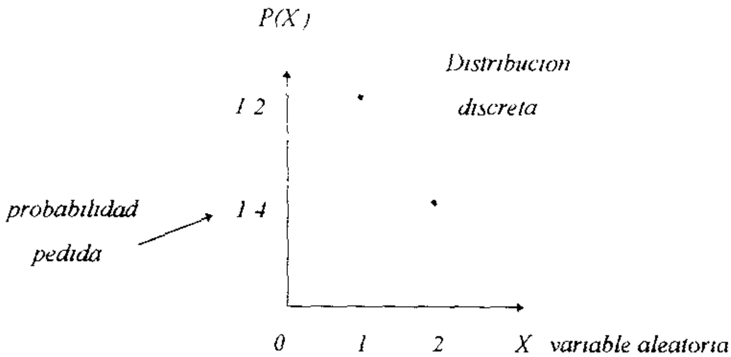
$$P(X = 2) = {}_2C_2 \; (1/2)^2 \, (1/2)^{2-2}$$

$$= \frac{2!}{2!(2-2)!} \, (1/4)$$

$= 1/4$ ***probabilidad pedida.***

Para poder representar en forma gráfica este problema, se debe tomar en cuenta la variable aleatoria de la que se este hablando, en este caso, X número de soles que caen al lanzar dos monedas. Tabulando y graficando

X	$P(X)$
0	$1/4$
1	$1/2$
2	$1/4$



de donde la media o valor esperado $E(X) = np = 2(\frac{1}{2}) = 1$; *la varianza* $\sigma^2 = npq = 2(\frac{1}{2})(\frac{1}{2}) = \frac{1}{2}$ *y la distribución estándar* $\sigma = \sqrt{npq} = \sqrt{2(\frac{1}{2})(\frac{1}{2})} = \frac{1}{2}$

EJEMPLO 5.3

Se estuvo observando a un policía practicar tiro al blanco durante un determinado tiempo y se vio que el promedio con el cual le pegaba era de 3/5. Si el policía tirará 5 veces ¿cuál sería la probabilidad de que le pegara por lo menos cuatro veces?.

Este problema puede ser fácilmente planteado por medio de la binomial en donde la probabilidad de éxito seria pegarle al blanco y ésta es de 3/5. Por lo que la probabilidad de fracaso que es igual ***f = 1 - p*** *= 1 - 3/5 = 2/5;* ***n*** *se definió como el número de veces en que es repetido el experimento (tambien llamado número de ensayos binomilaes) es igual a 5, la variable aleatoria que se estaría manejando sería aquella que representa el número de veces que el policía le pega al blanco. Como se esta pidiendo la probabilidad de por lo menos 4 aciertos, esto significa cuatro aciertos* *o más, por lo tanto será la suma de la probabilidad de exactamente cuatro aciertos más exactamente cinco. Sustituyendo los valores en la fórmula de la binomial y calculando las probabilidades deseadas, se tiene*

$$P(X = 4) = {}_5C_4\ (3/5)^4\,(2/5)^{5-4}$$

$$= \frac{5!}{4!(5-4)!}\left(\frac{81}{625}\right)\left(\frac{2}{5}\right)$$

$$= 5\,(.0518)$$

$$\boldsymbol{P(X = 4) = .2590}$$

así mismo,

$$P(X = 5) = {}_5C_5\,(3/5)^5\ (2/5)^{5-5}$$

$$= 1 \cdot (243/3125) \cdot 1$$

$$\boldsymbol{P(X = 5) = .0777}$$

ahora como

$$P(\text{por lo menos cuatro aciertos}) = P(X = 4) - P(X = 5)$$

$$= .2590 - .0777$$

$$= .3367$$

EJEMPLO 5.4

La probabilidad de que un estudiante que entra a la Escuela Nacional Colegio de Ciencias y Humanidades pase a la licenciatura es de .45 Obtener la probabilidad de que entre 5 estudiantes elegidos al azar:

i) al menos uno pase a la licenciatura.

ii) ninguno pase a la licenciatura.

Haciendo el planteamiento desde el punto de vista de éxito-fracaso, es decir, utilizando la binomial, se tiene.

p *(probabilidad de éxito) = .45*

q = 1 - p *(probabilidad de fracaso) = 1 - .45 = .55*

n *(número de pruebas repetidas) = 5*

X *(valores de la variable aleatoria) = 0,1,2,3,4,5*

NOTA. *Al menos uno, es una expresión que significa uno o más de uno.*

i) al menos uno pase la licenciatura.

La probabilidad pedida en este inciso es.

$$P(\text{al menos uno}) = P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5)$$

es decir, se deben calcular cinco binomiales

$$P(X=1) = {}_5C_1\,(.45)^1\,(.55)^{5-1}$$

$$= \frac{5!}{1!(5-1)!}\,(.45)(.0915)$$

$$= \frac{120}{24} (.0411)$$

$$= 5\,(.0411)$$

$$\boxed{P(X=1) = .2055}$$

$$P(X = 2) = {}_5C_2\,(.45)^2\ (.55)^{5-2}$$

$$= \frac{5!}{2!(5-2)!}\ (.2025)(.1663)$$

$$= \frac{120}{12} (.0336)$$

$$= 10\,(.0336)$$

$$\boxed{P(X=2) = .3367}$$

$$P(X = 3) = {}_5C_3\,(.45)^3\ (.55)^{5-3}$$

$$= \frac{5!}{3!(5-3)!}\ (.0911)(.3025)$$

$$= \frac{120}{12} (.0275)$$

$$= 10\,(.0275)$$

$$\boxed{P(X=3) = .2755}$$

$$P(X = 4) = {}_5C_4\,(.45)^4\ (.55)^{5-4}$$

$$= \frac{5!}{4!(5-4)!}\,(.0410)(.55)$$

$$= \frac{120}{24}\,(.0225)$$

$$= 5\,(.0225)$$

$$\boxed{P(X=4) = .1127}$$

$$P(X = 5) = {}_5C_5\,(.45)^5\ (.55)^{5-5}$$

$$= \frac{5!}{5!(5-5)!}\,(.0184)(1)$$

$$= \frac{5!}{5!}\,(.0225)$$

$$\boxed{P(X=1) = .0184}$$

por lo tanto

$$P(\text{al menos uno}) = P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5)$$

$$= .2055 + .3367 + .2755 + .1127 + .0184$$

$$= \mathbf{.9488}$$

ii) ninguno pase a la licenciatura.

$$P(X=0) = {}_5C_0\,(.45)^0\;(.55)^{5-0}$$

$$= \frac{5!}{0!(5-0)!}\,(1)(.0503)$$

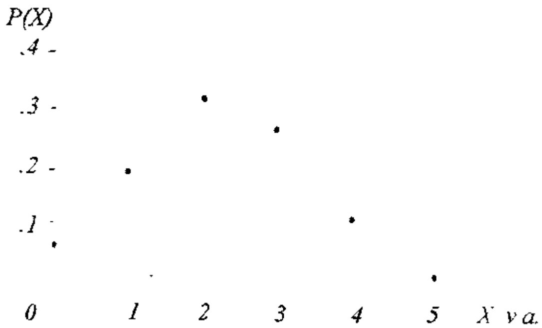
$$= (1)(1)\,(.0503)$$

$$\boxed{P(X=0) = .0503}$$

Ahora como X = al número de alumnos que logren llegar a la licenciatura.

X	P(X)
0	.0503
1	.2055
2	.3367
3	.2755
4	.1127
5	.0184

graficando.



5.3 ESTUDIO DE LA DISTRIBUCIÓN NORMAL COMO EJEMPLO DE FUNCIÓN DE DISTRIBUCIÓN CONTÍNUA.
Cálculo, Graficación e Interpretación.

Con anterioridad ya se mencionaron algunas características de la Distribución Normal o de Gauss. Sin embargo, no obstante que en este apartado se repitan algunos conceptos, se hace con el firme propósito de profundizar en su estudio.

La Distribución Normal es por muchos motivos la distribución de probabilidad más importante, y esto se debe a que:

- *Muchos problemas estadísticos reales se presentan en forma normal o muy semejante a ésta.*
- *Aún cuando su fórmula original es complicada y el cálculo de ésta, sería a base de integrales, al estandarizarla o tipificarla, con una sencilla fórmula (que más adelante será estudiada con detenimiento), resulta muy fácil obtener buenos resultados.*
- *Algunos casos discretos en los cuales es complicado su cálculo, la Distribución Normal proporciona una excelente aproximación.*

y se define como:

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}(X-\mu/\sigma)^2}$$

donde.

σ *Desviación Estándar*

σ^2 *Varianza*

$\pi = 3.1415\ldots$

μ *media*

$e = 2.718281\ldots$

X *Valores de la Variables Aleatoria*

en términos probabilísticos , se dice que bajo esta fórmula X variable aleatoria esta distribuida normalmente con media μ *y varianza* σ^2 *, que en símbolos* $X \sim N(\mu, \sigma^2)$ *. Sin embargo, esta fórmula no será utilizada en este trabajo, en su lugar, se usará la fórmula de la Normal Estandarizada..*

La forma como se estandiriza una Normal, es la siguiente:

Si X es una Variable Aleatoria con distribución normal, obtenida de un cierto experimento, será condición mínima y necesaria para estandarizarla, que a cualquier valor de la variable aleatoria **X** *se le reste el valor de la media aritmética* μ *y se divida entre la desviación estándar* σ*. Lo que se denota con la letra* **z.**

$$z = \frac{X - \mu}{\sigma}$$

El siguiente esquema muestra la forma en como una normal puede ser estandarizada.

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}(X-\mu/\sigma)^2}$$

Distribución Normal
$X \sim N(\mu, \sigma^2)$

$$z = \frac{X - \mu}{\sigma}$$

condición necesaria para estandarizar

$$f(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2}$$

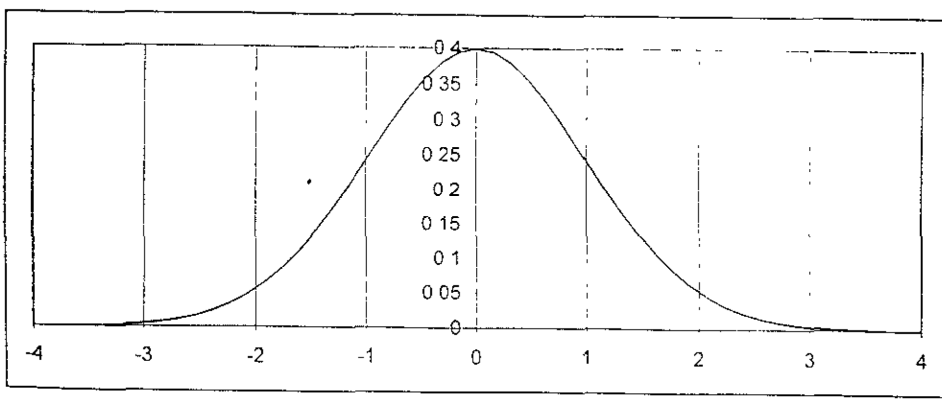
Normal Estandarizada
$z \sim N(0, 1)$

Una variable aleatoria que tenga una distribución normal estandarizada tendrá media $\mu = 0$ y desviación estándar $\sigma = 1$ y resulta fácil graficarla, para lo cual se tabulará en primera instancia y posteriormente se graficará

Se tiene:

X	y
4	0001
3	.0044
2	0539
1	.2419
0	.3989
-1	.2419
-2	.0539
-3	.0044
-4	.0001

de donde.



Es importante aclarar que cualquier probabilidad que se obtenga con la ayuda de la normal estandarizada, será igual a el área bajo la curva que esta entre z = 0 y cualquier valor estandarizado de X (siendo X una variable aleatoria cualquiera) Otra cosa que se debe mencionar, es que el valor total del área bajo la curva estandarizada es igual a 1, y como ésta es simétrica, tanto la mitad de el lado positivo como la de el lado negativo, tendrán valor de ½ . Así mismo, se puede fácilmente ver que cuando

$z > 3$ el área es casi igual a ½, por lo que cualquier valor de z 3 9 será considerado igual a ½. Por ultimo, en la tabla de la normal estandarizada solo marcan valores de z para el lado positivo pues cualquier valor negativo de z tendra el mismo valor que los que toma en el lado positivo.

En resumen la curva Normal Estándar cumple las siguientes propiedades:

- Es simétrica con respecto al eje de las y.

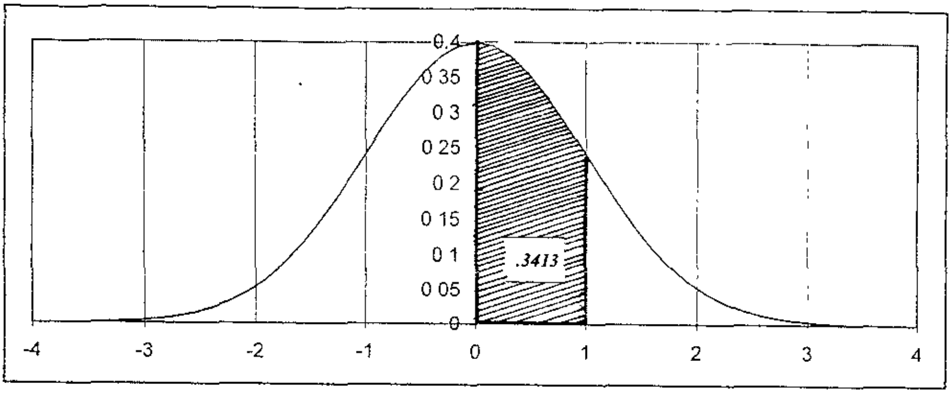
- La curva se aproxima al eje de las X más y más estrechamente cuando z es mayor que el valor absoluto ± 3. Sin embargo, la curva se extiende sin límite en ambas direcciones, puesto que $f(z)$ nunca es cero para cualquier valor de z.

- La curva tiene un punto máximo cuando $z = 0$.

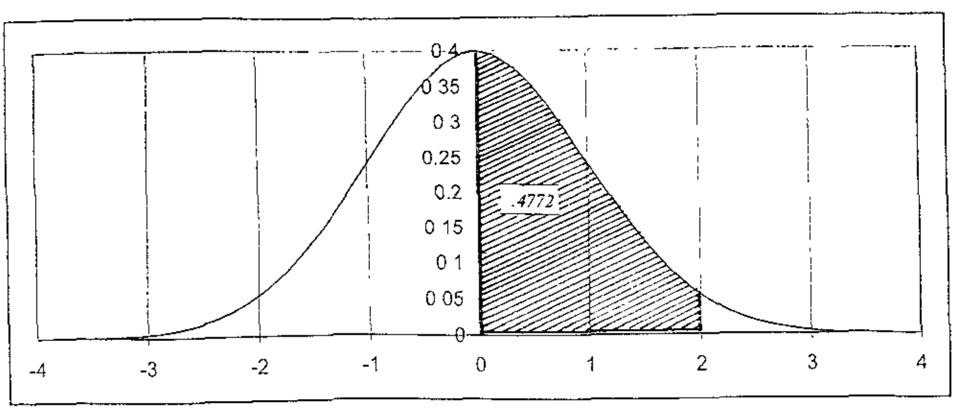
- Cualquier punto sobre la curva no dará la probabilidad de un evento, ésta la proporcionará el área bajo la curva.

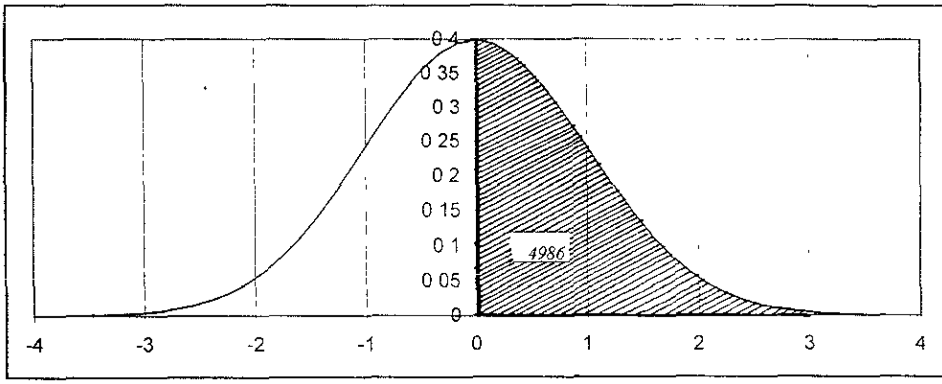
- El área total bajo la curva es igual a 1, y ésta se puede sacar fácilmente en las tablas del área bajo la curva normal estandarizada que aparece al final de estos apuntes. Así mismo, aparece un breve explicación de la forma en como son utilizadas.

Se puede ver en las tablas del área bajo la curva normal estandarizada que de 0 a donde $X = 1$, el valor en tablas es de .3413. De 0 a cuando $X = 2$, vale .4772 y de 0 a cuando $X = 3$, $.4986 \approx .5$.



El área sombreada en la gráfica anterior, muestra el área que hay entre $z = 0$ y $z = 1$ y fue tomada de las tablas de la normal estandarizada que se encuentran al final de este trabajo. De igual forma, el área bajo la curva de $z = 0$ a $z = 2$ esta representada en la siguiente gráfica.





Por último la tercera gráfica marca el área de z = 0 a z =3.

Estos valores representan en si, ya un valor de probabilidad, y se mostrará más claramente con un ejemplo.

NOTA : *Siempre que se quiera resolver un problema con la Normal estadarizada , se deben de tener u obtener, la media aritmética μ y la desviación estándar σ.*

EJEMPLO 5.5

En la Compañía www , se encuentran laborando un total de 700 obreros calificados , por lo que el dueño de dicha compañía, mando a hacer un estudio estadístico para contestarse algunas preguntas. Se obtuvieron los siguientes datos.

El ingreso mensual promedio de los 700 obreros es de \$ 500 00 y la desviación estándar es de \$ 100.00 así mismo, la población conserva un comportamiento normal. Las preguntas a contestar son, Si se toma un obrero al azar. ¿Cuál seria la probabilidad de que ganara \$ 600.00? y ¿cuántos trabajadores tienen ese ingreso?.

Para contestar estas preguntas, sabiendo que es un problema de normal, se debe estandarizar en primera instancia, aquí los datos que se tienen son los siguientes

$$\mu = \$\ 500.00$$

$$\sigma = \$\ 100.00$$

$$X = \$\ 600.00$$

estandarizando.

$$z = \frac{X - \mu}{\sigma}$$

$$= \frac{600 - 500}{100}$$

$$= \frac{100}{100}$$

$$z = 1$$

por lo que buscando en tablas del área bajo la curva normal estandarizada, el área que hay de 0 a 1, es de .3413, o lo que es lo mismo, la probabilidad de que un obrero tomado al azar gane \$ 600.00 , es de .3414. En símbolos .

$$p\ (X = 600) = .3413$$

por lo que el 34.13% de los 700 obreros ganan \$ 600.00, bastaría con calcular el 34.13% de 700, lo que da.

No. de obreros que ganan \$ 600.00 = 700 x 34.13%

$$= 238.91$$

$$\approx 239 \text{ obreros.}$$

Las preguntas más o menos comunes en lo que se refiere a la normal son

- *Cuando se pregunta sobre un valor puntual de la variable aleatoria.* $P(X = x)$
- *Cuando se pregunta sobre un valor mayor o mayor o igual al valor de un punto de la variable aleatoria.* $P(X > x)$ o $P(X \geq x)$.
- *Cuando se pregunta por un valor menor o menor o igual a un punto de la variable aleatoria.* $P(X < x)$ o $P(X \leq x)$.
- *Cuando se pregunta sobre un intervalo.* $P(x_0 \leq X \leq x_1)$.

A continuación se ejemplificarán todos los puntos anteriores en un ejercicio

EJEMPLO 5.6

Una empresa que fabrica tornillos, tubo una producción de 10,000 tornillos. Bajo un estudio, se observó que los diámetros de los tornillos se comportaban de una manera normal con media .25 pulgadas y desviación estándar de .02 pulgadas. Se considera defectuoso un tornillo si su diámetro es menor que .20 pulgadas o mayor que .28. Obtener:

i) el porcentaje de tornillos defectuosos y el monto.

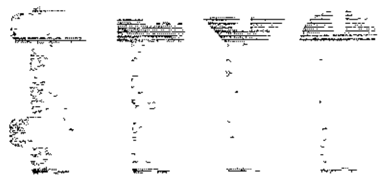
ii) la probabilidad de que tomado un tornillo al azar tenga diámetro igual a .27 pulgadas.

iii) entre .23 y .27 pulgadas.

iv) entre .26 y .29

En este ejercicio, $\mu = .25$ *pulg. Y* $\sigma = .02$ *pulg*

ı) el porcentaje de tornillos defectuosos y el monto.



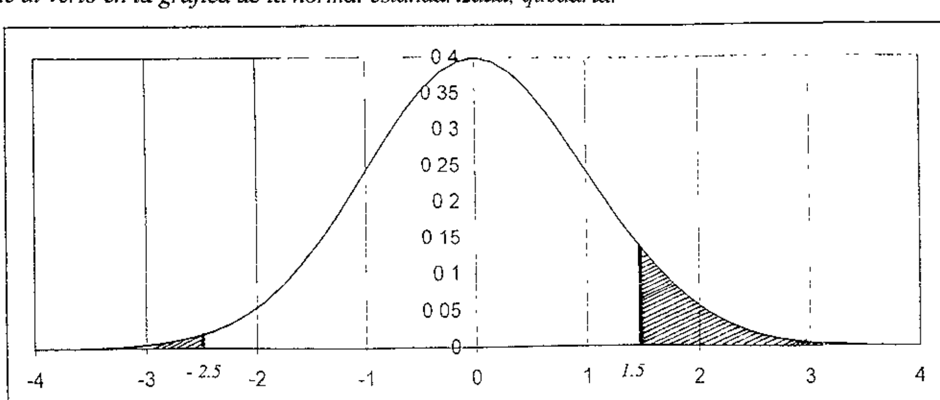
como un tornillo se considera defectuoso si su diámetro es menor que .20 pulgs o mayor que .28 pulgs. Lo primero que se debe hacer es estandarizar los dos valores anteriores.

$$z_1 = \frac{.20 - .25}{.02} = -2.5$$

y

$$z_2 = \frac{.28 - .25}{.02} = 1.5$$

que al verlo en la gráfica de la normal estandarizada, quedaría.



y buscando en tablas

$$P(X < .20) = 4938$$

sin embargo, esta probabilidad es de tornillos buenos que cumplen la norma permitida y lo que se desea obtener es tornillos que no la cumplen. Se sabe que el área total de cualquiera de las dos mitades de la normal estandarizada vale ½ , por lo que al restarle la cantidad obtenida se encontrará el área deseada, es decir.

$$\begin{array}{r} .5 \\ -\ .4938 \\ \hline .0062 \end{array}$$

que sería la probabilidad de tornillos fabricados que tienen diámetro menor que lo permitido, y que dado en porcentajes, esto sería igual a .62% . Bastará multiplicar el total de tornillos fabricados por el porcentaje obtenido.

10,000 x .62% = 62 tornillos tienen diámetro menor que lo permitido

por otro lado.

$$P(X > .28) = .4338$$

que de igual forma.

$$\begin{array}{r} .5 \\ -\ .4338 \\ \hline .0662 \end{array}$$

que sería la probabilidad de tornillos con diámetros más grandes de lo permitido y por lo tanto defectuosos. El porcentaje es 6.62% que al multiplicar por la cantidad de tornillos fabricados quedaría.

$$\begin{array}{r} 10{,}000 \\ \underline{\times\ 6.62\%} \\ 662 \end{array} \quad \textit{tornillos.}$$

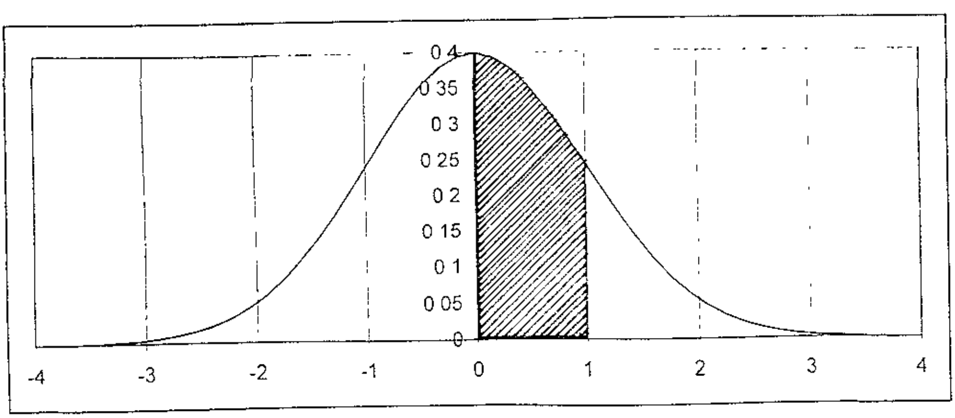
Por lo tanto el porcentaje de tornillos defectuosos es 7.24% y el número de tornillos defectuosos es 724.

ii) La probabilidad de que tomado un tornillo al azar tenga diámetro igual a 27 pulgs.

estandarizando

$$z = \frac{.27 - .25}{.02} = 1$$

gráficando



buscando en tablas

$$P(X = .27) = .3413 \text{ probabilidad pedida.}$$

iii) entre .23 y .27 pulgs

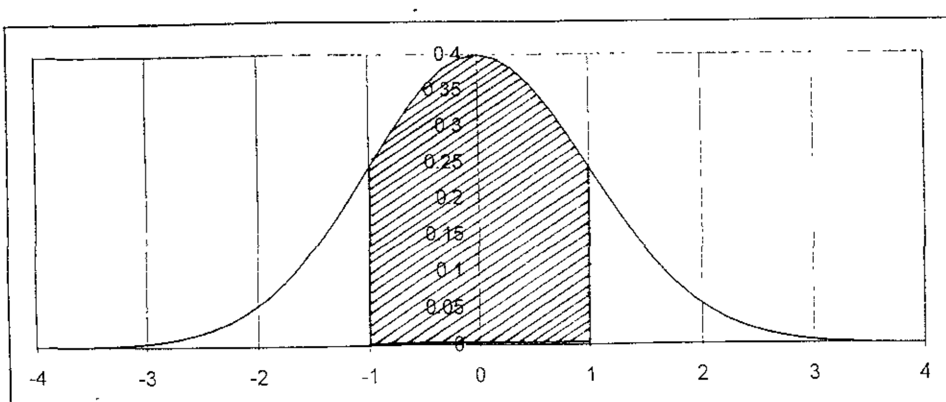
en este caso, se esta preguntando la probabilidad de un intervalo, que poniéndolo en terminos de probabilidad se expresa como $P(.23 \leq X \geq .27)$ *por lo que se estandarizaran los extremos.*

$$z_1 = \frac{.23 - .25}{.02} = -1$$

y

$$z_2 = \frac{.27 - .25}{02} = 1$$

que al visualizarlo en la gráfica de la normal estandarzada, quedaria.



y buscando en tablas.

Valor estandarizado de X	*valor en tablas*
- 1	*.3413*
1	*.3413*

y deberán ser sumados los valores en tablas para encontrar la probabilidad pedida.

$$\begin{array}{r} .3413 \\ +\ .3413 \\ \hline .6826 \end{array}$$

por lo tanto

$$P(.23 \leq X \geq .27) = .6826$$

iv) entre .26 y .29

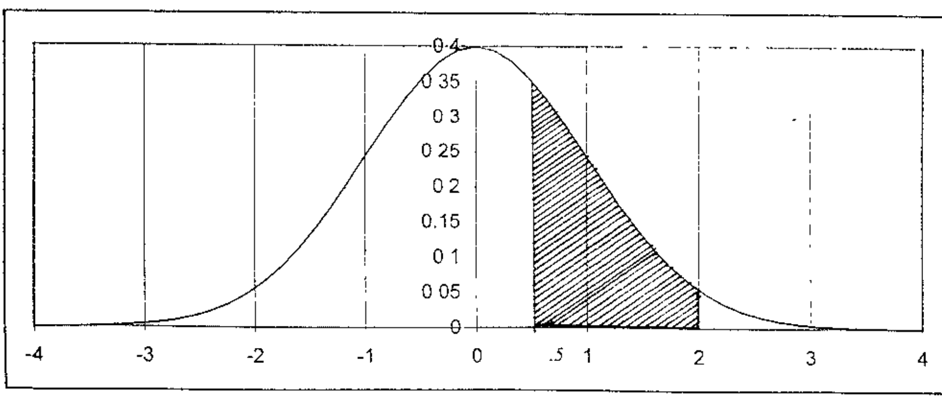
Lo mismo que en le caso anterior, se esta buscando la probabilidad de un intervalo, por lo que se calculará $P(.26 \leq X \geq .29)$. *Estandarizando ambos extremos.*

$$z_1 = \frac{.26 - .25}{.02} = .5$$

y

$$z_2 = \frac{.29 - .25}{.02} = 2$$

que al visualizarlo en la gráfica de la normal estandarzada, quedaría.



y buscando en tablas.

Valor estandarizado de X	*valor en tablas*
.5	*.1915*
2	*4772*

para obtener la probabilidad pedida, se debe restar al valor mayor en tablas, el valor menor.

6.- CONCLUSIONES Y RECOMENDACIONES.

Durante el tiempo de elaboración de estos apuntes, se tuvieron grandes experiencias, ya que varios fueron los comentarios , que por parte de los profesores del Colegio de Ciencias y Humanidades, enriquecieron su contenido, así mismo, se trató de visualizar diferentes criterios docentes sobre un sólo tema y consultar un número más o menos considerable de fuentes.

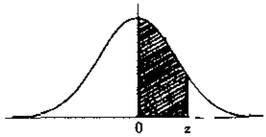
Es importante antes de vertir las conclusiones y recomendaciones de este trabajo, mencionar que el nuevo programa de Probabilidad y Estadística I es muy ambicioso al ser comparado con el viejo programa, otro detalle de suma importancia es, que el nuevo programa tendrá que ser aplicado varias veces, según mi punto vista, para poder hacer comentarios sobre la verdadera problemática con la cual el profesor y el alumno tendrán que enfrentarse , así como, para unificar los criterios de profundidad.

Muchos cursos y comentarios previos a la implantación de éstos se han hecho, sin embargo insisto, sólo la experiencia y el tiempo podrán proporcionar el exito o fracaso de estos nuevos programas.

Dicho lo anterior, las conclusiones se harán después de haber aplicado por primera vez dicho programa y estas son las siguientes:

- *El programa tendrá que ser adaptado al tiempo que se tiene para ser impartido, sin olvidar que el aprendizaje y el dominio de los temas es el principal objetivo*
- *Con respecto a los antecedentes, el alumno que se enfrenta a la materia, carece de bases en relación a la teoría de conjuntos, que es de suma importancia para tener éxito en esta asignatura.*
- *Se tendrán que realizar más cursos para la unificación de criterios en lo que se refiere a la profundidad y al nivel en el cual deberán de ser tratados los temas*
- *Por último, es recomendable elaborar lo más pronto posible, toda la nomenclatura de apoyo, como son, guías de estudio, folletos, libros, etcétera, que apoyen a los alumnos para lograr éxito y si estos ya existen difundirlos.*

Areas bajo la Curva Estándar normal de 0 a z



z	0	1	2	3	4	5	6	7	8	9
0.0	.0000	0040	.0080	.0120	.0160	.0199	.0239	0279	0319	0359
0.1	.0398	.0438	.0478	0517	.0557	0596	.0636	0675	.0714	0753
0 2	.0793	.0832	.0871	.0910	.0948	.0987	1026	1064	.1103	1141
0.3	.1179	.1217	.1255	1293	.1331	.1368	.1406	1443	1480	1517
0.4	.1554	.1591	.1628	.1664	.1700	1736	.1772	1808	1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	2157	.2190	.2224
0.6	.2258	.2291	.2324	.2357	2389	2422	.2454	.2486	2518	.2549
0.7	.2580	.2612	.2642	2673	2704	.2734	.2764	2794	2823	.2852
0.8	.2881	.2910	.2939	.2967	.2996	3023	.3051	.3078	.3106	3133
0 9	.3159	.3186	.3212	.3238	.3264	.3289	3315	3340	3365	.3389
1.0	.3413	.3438	3461	.3485	3508	.3531	.3554	3577	3599	3621
1.1	.3643	.3665	.3686	.3708	.3729	3749	.3770	.3790	3810	3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	3962	3980	3997	.4015
1.3	.4032	.4049	.4066	.4082	4099	4115	4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	4251	4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	4394	.4406	.4418	4429	4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	4525	4535	.4545
1.7	.4554	.4564	4573	.4582	4591	.4599	.4608	4616	4625	.4633
1.8	.4641	.4649	.4656	4664	.4671	4678	.4686	4693	4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	4744	.4750	4756	4761	.4767
2 0	.4772	.4778	.4783	.4788	.4793	4798	4803	.4808	4812	.4817
2.1	.4821	.4826	4830	4834	.4838	.4842	4846	.4850	.4854	.4857
2.2	.4861	4864	.4868	.4871	.4875	4878	4881	4884	4887	.4890
2 3	.4893	.4896	.4898	.4901	4904	4906	.4909	4911	.4913	.4916
2.4	.4918	.4920	.4922	4925	.4927	.4929	.4931	4932	4934	.4936
2 5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	4951	.4952
2.6	.4953	.4955	.4956	.4957	4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	4971	.4972	.4973	.4974
2 8	.4974	.4975	.4976	.4977	.4977	.4978	4979	.4979	4980	.4981
2 9	.4981	.4982	4982	.4983	.4984	4984	4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	4989	.4989	4989	4990	.4990
3.1	4990	.4991	4991	.4991	4992	.4992	.4992	.4992	4993	.4993
3 2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	4996	.4996	4996	4996	.4997
3 4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	4997	4997	4998
3 5	.4998	.4998	.4998	4998	.4998	4998	4998	4998	4998	.4998
3 6	.4998	.4998	4999	.4999	4999	.4999	4999	.4999	4999	.4999
3.7	.4999	.4999	.4999	4999	.4999	4999	.4999	.4999	4999	.4999
3 8	.4999	.4999	.4999	4999	.4999	.4999	4999	4999	4999	.4999
3.9	.5000	.5000	.5000	5000	.5000	.5000	5000	5000	5000	.5000

USO DE LAS TABLAS DE LA NORMAL ESTANDARIZADA

La forma en como son usadas las tablas de la normal estandarizada es muy semejante al uso de las tablas para sacar logaritmos (dicho esto, para aquellas personas que con anterioridad hayan hecho uso de dichas tablas para obtener logaritmos). Sin embargo, aprender a utilizar la tabla de la normal es sencillo. Primero se debe saber que hay varios tipos de tablas para conocer el área bajo la curva normal estandarizada, no obstante, en este trabajo, se tomó la decisión de tomar este tipo, por considerar que su uso es relativamente fácil.

Hay que hacer notar que esta tabla es para cuando los valores de z toman dos decimales, y que éstos están representados en la columna de la izquierda con un sólo decimal, dejando en el renglón de z, que se encuentra en la parte superior, un decimal más que completa los dos decimales mencionados. Dicho renglón, en lo sucesivo, será llamado de números complementarios. Por ejemplo, si al estandarizar un valor de X (variable aleatoria), diera un como resultado 0.55, se busca en la columna de z el valor de 0.5 y en el renglón de números complementarios el número 5, completando así, el valor pedido 0.55 dando como resultado .2088. Tomando otro ejemplo, supóngase que se quiere encontrar el valor ya estandarizado de 1.15, se localizará 1.1 en la columna de z y el número 5 en el renglón de números complementarios, y su valor en tabla será .3749

BIBLIOGRAFÍA

- *Guerra, Gómez, Rodriguez Roberto A. "El CCH una experiencia universitaria" México UNAM.*
- *Asociación Nacional de Universidades e Instituciones de Enseñanza Superior, México, 1983.*
- *Mood and Graybill. Introduction to The Theory of Statistics. Ed. McGraw-Hill, Second Edition.*
- *Erwin Kreyszig. Introducción a la Estadística Matemática Principios y Métodos. Ed. Limusa*
- *Stephen P. Shao. Estadística para Economistas y Administradores de Empresas. Ed. Herrero Hermanos, Sucs., S. A. México.*
- *Juana Castillo Padilla. Estadística Inferencial Básica. Colección 90.2. Colegio de Ciencias y Humanidades, UNAM.*
- *Robert Johnson. Estadística Elemental. Ed. Grupo Editorial Iberoamericana.*
- *William Feller Introducción a la Teoria de Probabilidades y sus Aplicaciones, Vol. I Ed. Limusa.*
- *Murray R. Spiegel. Estadística. Libros McGraw-Hill, Serie compendios Schaum.*
- *Seymour Lipschuts. Probabilidad. Ed. McGraw-Hill.*